NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

DESIGN OF OPTIMAL CYCLERS USING SOLAR SAILS

by

Robert E. Stevens

December 2002

Thesis Advisor:	I. Michael Ross
Second Reader:	Dennis Byrnes

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** December 2002	**3. REPORT TYPE AND DATES COVERED** Aero and Astronautical Engineer Degree	
4. TITLE AND SUBTITLE. Optimal Cycling Trajectories Using Solar Sails			**5. FUNDING NUMBERS**
6. AUTHOR(S) Robert Stevens			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT ***(maximum 200 words)***

Ongoing interest in establishing a base on Mars has spurred a need for regular and repeated visits to the red planet using a cycling shuttle to transport supplies, equipment and to retrieve surface samples. This thesis presents an approach to designing an optimal heliocentric cycling orbit, or cycler, using solar sails. Results show that solar sails can be used to significantly reduce V_∞s at Mars and Earth. For example, using a reasonably high performance solar sail, a $V_\infty|_{Mars} = 2.5$ km/s is possible at every synodic period using a two-dimensional orbit model. Lower performance sails were also modeled resulting in paths that behaved more like a ballistic Aldrin cycler with higher V_∞s. Double rendezvous missions were explored where the spacecraft must match the velocities of both Earth and Mars, offering promising trajectories for Mars sample return missions. The solutions to these missions, although not necessarily cyclers, show that using a sail to rendezvous with and remain near Mars for an optimal amount of time will minimize the total transit time between Earth and Mars. General-purpose dynamic optimization software, DIDO, is used to solve the optimal control problem using a pseudospectral method using both two- and three-dimensional elliptic orbit models.

14. SUBJECT TERMS Cycler Trajectory, Solar Sail, Optimal Design, DIDO, Mars Mission, Trajectory Design, Cycling Orbits			**15. NUMBER OF PAGES** 141
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

DESIGN OF OPTIMAL CYCLERS USING SOLAR SAILS

Robert E. Stevens
Lieutenant Commander, United States Navy
B.S., United States Naval Academy, 1989

Submitted in partial fulfillment of the
requirements for the degree of

**AERONAUTICAL AND
ASTRONAUTICAL ENGINEER**

from the

**NAVAL POSTGRADUATE SCHOOL
December 2002**

Author: Robert E. Stevens

Approved by: I. Michael Ross
Thesis Advisor

Dennis Byrnes
Second Reader

Max F. Platzer
Chairman, Department of Aeronautics and Astronautics

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Ongoing interest in establishing a base on Mars has spurred a need for regular and repeated visits to the red planet using a cycling shuttle to transport supplies, equipment and to retrieve surface samples. This thesis presents an approach to designing an optimal heliocentric cycling orbit, or cycler, using solar sails. Results show that solar sails can be used to significantly reduce V_∞s at Mars and Earth. For example, using a reasonably high performance solar sail, a $V_\infty|_{Mars} = 2.5$ km/s is possible at every synodic period using a two-dimensional orbit model. Lower performance sails were also modeled resulting in paths that behaved more like a ballistic Aldrin cycler with higher V_∞s. Double rendezvous missions were explored where the spacecraft must match the velocities of both Earth and Mars, offering promising trajectories for Mars sample return missions. The solutions to these missions, although not necessarily cyclers, show that using a sail to rendezvous with and remain near Mars for an optimal amount of time will minimize the total transit time between Earth and Mars. General-purpose dynamic optimization software, DIDO, is used to solve the optimal control problem using a pseudospectral method using both two- and three-dimensional elliptic orbit models.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF SYMBOLS

Subscripts:

$0, i, f$ - Initial, intermediate, or final conditions respectively

u, l - Upper or Lower bounds respectively

e, m, p - Earth, Mars or planet respectively

E - Heliocentric-ecliptic frame

Superscripts:

$-, +$ - Immediately prior to or following an interior event respectively

$*$ - Optimal

Variables:

Dots denote the time derivative of a given variable.

$\wedge$ denotes a unit vector direction.

a - Semi-major axis

E - Mayer Cost Functional or eccentric anomaly

e - Eccentricity

$\mathbf{e}$ - Events function

F - Lagrange Cost Functional

$\mathbf{f}$ - Right-hand side of Dynamic equations

$\mathbf{g}$ - Path function

H - Hamiltonian

$\mathbf{h}$ - Orbital Angular Momentum vector

h - Magnitude of $\mathbf{h}$

i - Orbital inclination

J - Scalar cost function or performance measure

L - Lagrangian or augmented function

M - Mean anomaly

m - Mass of spacecraft

N - Number of nodes

N_D - Number of dimensions

N_k - Number of interior knots

n - Mean motion of planet

$\mathbf{p}$ - Vector of parameters

p - Semi-latus rectum

$\mathbf{r}$ - Radius vector from the sun

r - Magnitude of $\mathbf{r}$

S - Arbitrary bounding set for states

$\mathbf{T}$ - Thrust vector

T - Magnitude of $\mathbf{T}$

t - Time

U - Arbitrary bounding set for controls

$\mathbf{u}$ - Control vector

u - Radial velocity

V_∞ - Hyperbolic excess speed

v - Transverse or along-track velocity

$\mathbf{x}$ - State vector for spacecraft

a - Solar sail cone angle

b - Solar sail lightness number

d - Solar sail clock angle or gravity assist deflection angle

f - 3rd spacecraft state variable

g - Weighting factor in convex combination cost function

λ - Lagrange multipliers for dynamic constraints

μ - Gravitational parameter of sun

m_a, m_d - Control covectors for cone and clock angles respectively

q - Spacecraft angular displacement in ecliptic

? - Angular velocity vector

ω - Argument of Perigee

Ω - Right Ascension of the Ascending Node

$\mathbb{R}^{N_D}$ - N_D-dimensional real space

z - Earth-Mars lead angle

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

Thanks to…

The Lord God from whom all blessings flow.

My wife, Jodi, and children, Luke and Kate for their encouragement and untiring devotion. I love you all very much.

My thesis advisor, Mike Ross for his enthusiastic guidance and instruction throughout this whole project.

Dennis Byrnes and Jon Sims at the Jet Propulsion Lab who supplied valuable insights into cycler trajectory modeling and design.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Spacecraft orbiting along cycling trajectories between Earth and a nearby planet such as Mars could be tremendously valuable for missions ranging from sample returns to ferrying supplies and personnel between the planets like escalators. Naturally occurring cycler trajectories would be ideal, requiring no added energy other than that provided by the gravity fields of the target planets. If such "free ride" paths exist, they could host spacecraft making an infinite number of round trips without requiring propulsion systems. This hypothetical path would be characterized by a purely natural Keplerian trajectory that repeats itself endlessly interrupted only by planetary swingbys. Such cyclers could be analyzed for various target planets, however Earth-Mars cyclers will be studied in this thesis because of relevance to establishing a base on our neighboring red planet. To this end, two prominent Earth-Mars cycler concepts have been proposed- the Aldrin cycler[1] and VISIT cycler[2]. The Aldrin cycler uses gravity assists from both planets and small well-timed ΔVs to maintain a continuous cycling trajectory. While the revisit times are appealing (7 round trips in 15 years), on-board propellant is required to provide a 15 year cumulative ΔV of approximately 2 km/s. Because no fuel is expended trying to reduce speed near a planet, hyperbolic excess speeds are high, in excess of 5 km/s. A VISIT cycler, on the other hand, uses neither gravity assists nor fuel for ΔV burns for orbit shaping, but rather resides in a natural heliocentric orbit that makes regular passes of both Earth and Mars. The advantage to this cycler is that no propellant is required to maintain the orbit for up to 20 years, but the primary disadvantages are the rather long revisit times (3 Earth visits and 4 Mars visits in 15 years) and large approach distances needed to avoid unwanted gravitational interaction. The purpose of this paper is to present an Earth-Mars cycler concept using solar sails that has the advantages of both cyclers without the on-board propellant.

Because it has been shown that a nearly ballistic cycler orbit can be maintained for 15 years[1], it is reasonable to speculate that a solar sail cycler is not only feasible, but more capable since it provides free controls. With solar radiation pressure providing free thrust to the solar sail, a cycler orbit may be maintained with short revisit times as well as

slow planetary approach speeds. Small approach velocities to Earth and Mars at short distances are desirable for operational reasons and because they exhibit attractive orbit-shaping characteristics due to large angle swingby deflections.

Given that a solar sail requires no propellant, we are free to determine a cost function that does not minimize the fuel. What then defines an optimal solar sail cycle r? Often the time of flight or an undesirable cycler characteristic like large approach velocities may be minimized. A design parameter of the sail itself, such as the size of the sail may be the set cost function. This thesis presents a framework for solving such optimal low-thrust cycler trajectories using the general-purpose dynamic optimization software DIDO[3]. It is worth mentioning that although solar sails are chosen as the propulsion method and cycling orbits are selected as the mission for this thesis, the framework presented here is applicable to any type of propulsion system and mission.

II. APPROACH TO OPTIMAL CONTROL SOLUTIONS

A. ROADMAP

At the outset of this thesis work, there were no known works related to optimal Earth-Mars solar sail cyclers, therefore a cautious approach to the optimal control solution was necessary. Starting with "simple" two point boundary value astrodynamic problems, a systematic approach was used that gradually increased the model complexity to arrive at solutions. These solutions were used to both validate the numerical optimization method and to obtain milestone profiles providing initial guesses and bounds to more complex problems. Benchmark optimal control problems were solved using DIDO, and the output solutions were compared with known solutions. Benchmark problems included Mars flyby, Mars rendezvous and Earth-Mars-Earth double rendezvous missions. Numerical results were also propagated using initial conditions and the output control histories for further validation. For many of the simpler problems (i.e. no intermediate event conditions) a costate profile was obtained corresponding to all the Lagrange multipliers at certain points in time along the path. Using the costates with the derived solar sail steering control law enabled a check of compatibility between the DIDO costate and output control histories. The resulting Hamiltonian output was also available to check for optimality. After the benchmark problem solutions had been verified, additional constraints and event conditions were included to obtain the more complex cycler trajectory that includes planetary swingbys of Earth and Mars.

The above approach was used for both two-dimensional and three-dimensional models. The two-dimensional solutions approximated both Earth and Mars orbits as being circular and coplanar serving as an initial estimate for the higher fidelity models. Fortunately, the relative inclination of Mars' orbit with respect to the Earth ecliptic is only 1.85° and both orbit eccentricities are not too large, so this approximation is not too bad. Increasing the fidelity of the cycler model, Mars was given an elliptic, but coplanar orbit to observe the differences in the optimal paths. Finally, Earth and Mars orbits were inclined and both eccentricities were taken into account. The results from the battery of 2-D problems provided bounds and initial guesses for their 3-D counterpart problems.

Comparison of the 2-D and 3-D solutions to the same optimal control problem provided some interesting insights into astrodynamic optimization.

1. Coordinate Systems

To improve the performance of the numerical analysis, it served well to choose a coordinate system with states that change slowly with time and dynamics that have few or no singularities. Using a coordinate system in which states vary relatively slowly with time seemed to allow accurate solutions with fewer optimization "node" points, or points along the path that are used to approximate the continuous states and controls in the numeric optimization process. Additionally, singularity avoidance is necessary to obtain feasible solutions.

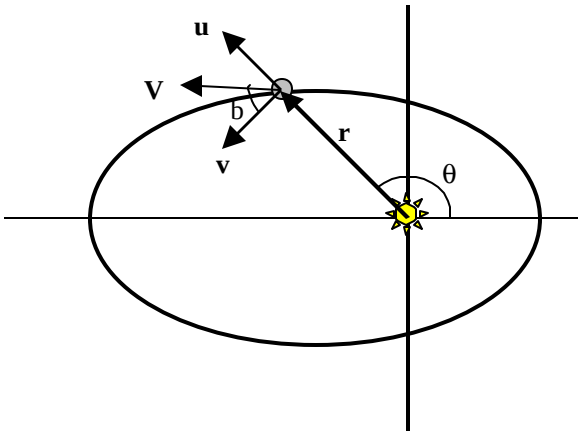

Figure 1 Elliptic Orbit in Polar Coordinates

For the two-dimensional model, a heliocentric polar coordinate system was used as shown in Figure 1. Because the maximum solar sail thrust is inversely proportional to the distance from the sun, this coordinate system was well suited for this problem. The only singularity in the polar equations of motion occurs at the center of the sun - an easily avoidable situation with proper state bounds (see Appendix A). Velocity components were expressed in terms of radial velocity and along-track velocity (ref 4 p. 43). The along-track velocity will be referred to as the *transverse* velocity and is defined as being normal to radial position vector in the orbital plane. Since the target orbits were approximated as circular and coplanar in the 2-D model, there was no need to define a principle direction (i.e. along vernal equinox).

Likewise, for the 3-D model spherical coordinates were chosen to represent 3-D states and dynamics because the resulting optimal state and control profiles were expected to vary slowly with time allowing excellent numerical approximations using relatively few nodes. Furthermore, these coordinates made it very simple to model the thrust due to the sail since thrust is dependent on radial distance, r (see Appendix B). By comparison, states of orbiting bodies in Cartesian coordinates oscillate. This could have lead to an undesirable condition in case there turned out to be too many cycles in the solution for the number of nodes to approximate accurately. It should be kept in mind, however, that Cartesian coordinates present perhaps the best choice when more perturbations are desired in the dynamics model (in this case, more nodes should be used). Equinoctial elements provide a set of singularity-free slowly changing variables, but because it was desirable to keep boundary conditions simple and "intuitive", they were not used (ref 4 p. 142).

2. Scaling

With the coordinate system chosen to reduce the number of optimization nodes required to accurately approximate states and controls, it became imperative to select units that would reduce numerical conditioning by avoiding the extremely large and small numbers. There is no need to confine the states to familiar units such as miles, kilometers, seconds, etc. A useful unit system for modeling interplanetary travel is the canonical system. Choosing a distance unit (DU) to be 1 A.U. and the normalized solar gravitational parameter μ to be 1, other state units may be calibrated accordingly. All states and controls have units that are of the same order of magnitude and dimensions are scaled by a factor of some characteristic dimension. This non-dimensionalization process also enhances visualizing metric relationships. For instance, the circular Earth orbit radius is 1 DU while the Mars orbit radius is 1.524 DUs, over 50% greater than Earth's. Heliocentric canonical units are summarized in Table 1.

Dimension	*Non-dimensional unit*	*mks equivalent*	*Comments*
Distance unit	DU	1.496E8 km	Semi-major axis of Earth orbit = 1 AU.
Time unit	TU	5.023E6 sec	Time to traverse 1 radian along circular orbit at 1 AU
Speed	VU = DU/TU	29.78 km/s	Velocity of body in 1 AU circular orbit.
Gravitational parameter	$\mu = 1\ DU^3/TU^2$	1.327E11 km^3/s^2	Solar gravitation parameter

Table 1 Canonical Units Used to Non-Dimensionalize the States and Parameters.

3. Bounding the Problems

Before approaching the full cycler model accounting for gravity assists, simpler two-point boundary value problems were analyzed. As already mentioned, this permits validation of the optimization software. Additionally, important insights may be gained by observing the behavior of simpler benchmark problems. For both the 2-D and 3-D models, three benchmark problems were solved prior to the coding of the cycler models. These are the Mars flyby, Mars rendezvous, and Earth-Mars-Earth double rendezvous round trip missions. The Mars flyby mission is a simple minimum time trajectory from Earth orbit to Mars orbit. This solution provides the absolute shortest time that a solar sail of a given performance can reach the Mars orbit. Other missions that include an Earth to Mars transit may use this time as a lower bound since it is impossible to reach Mars in less time. A Mars rendezvous mission is one in which the solar sail is tasked to reach Mars at its local orbital velocity ($V_\infty = 0$ with no Mars gravity acting) in minimum time. The resulting profile is not only useful in observing optimal rendezvous behavior, but also for bounding the double rendezvous problem that followed. The double rendezvous mission required the spacecraft to rendezvous with Mars, then track the planet for a time (with no effect due to Mars gravity), and then return to rendezvous with Earth. This mission introduced phasing constraints to ensure that Earth was there when the spacecraft returned. With the double rendezvous bounds established in this manner, a state discontinuity was introduced representing a gravity assist from the Mars encounter

making a zero-sphere patched conic cycler model complete. At this point, all the models and conditions had been established for a solar sail cycler and coded for use with the numeric optimization software, DIDO.

4. Numerical Analysis, terminology and DIDO [3]

Ideally, the solution to the optimal trajectory would view the entire space-time path of a spacecraft at once and simultaneously locate the optimal position for each point along that path meeting certain boundary conditions and dynamic constraints (Figure 2 and Figure 3). Limiting ourselves to current computer methods, however, we settle for using certain points along the space time path that will enable us to approximate the continuous path, thus discretizing the trajectory into a workable form for a non-linear programming (NLP) solver. An Euler method would provide an approximation that would be burdened by large errors. It was desired for this complex problem to use the best numerical approximations possible in solving the optimal control problem (OCP).

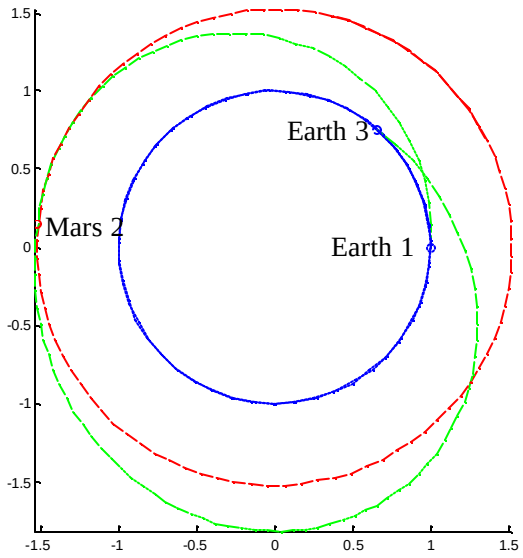

Figure 2 Path and Target manifolds (orbit positions)

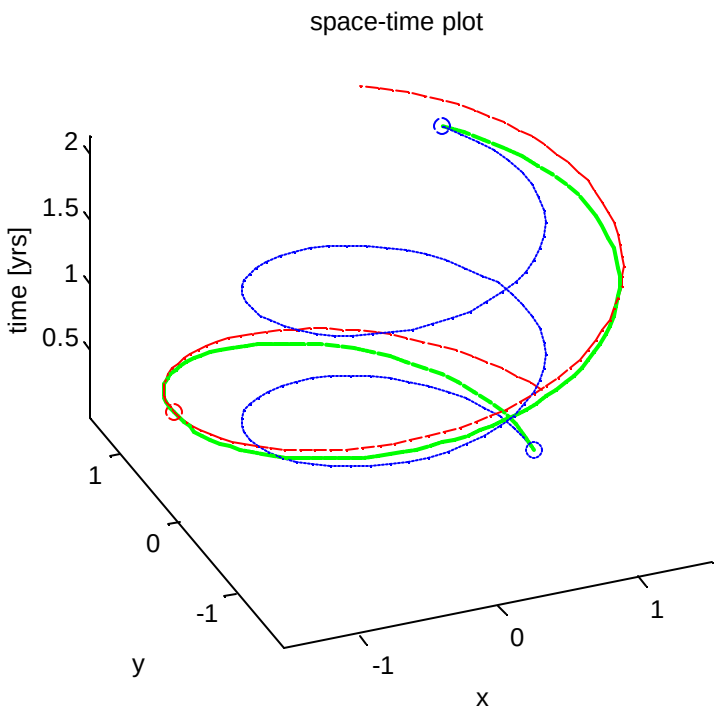

Figure 3 Path Meeting Event Manifolds (Planet Encounters) in Space (x-y) and Time.

The numerical solution to the optimal control problem presented in this thesis uses the software DIDO to interface with a generic third party NLP solver. The NLP solver requires an NLP variable vector, constraints and a cost function. DIDO provides these requirements set up in a numerically optimal fashion. All state and control path information are captured at *unevenly* distributed node points, thus the entire dynamic problem can be viewed as a static problem with the complete trajectory history at selected node points manifested as a single variable vector. The node points are selected in such a manner that the state and control approximation error is minimized. These numerically optimal node points are called Legendre-Gauss-Labatto (LGL) points and correspond to the zeros of the first derivative of the Nth degree Legendre polynomial[5,6]. The continuous states and controls are approximated by using the variables at the predetermined LGL node points and Lagrange interpolating polynomials, which are well suited for interpolation between unevenly distributed data points.

Dynamic constraints are imposed by forcing the derivatives of the approximated states to equal user-defined differential equations (equations of motion) at all the node points. Obtaining the derivative of the state variable vector is accomplished using a differentiation matrix defined in ref 5. This process of discretization and differentiation is called the Legendre pseudospectral method and has been used successfully in a variety of applications (refs 5-8).

Initial, intermediate and final event conditions are supplied by the user and structured into a boundary condition vector that DIDO provides as constraints to a third-party NLP solver. The approach used in this numerical solution to optimal control does not require propagation or shooting type methods. An initial guess is required, however it does not need to be a good one, or even a feasible one. Typically the end conditions are supplied and DIDO interpolates between these points as a guess.

To assist in the reading of the numerical methods used in this thesis, some definitions from ref 3 are supplied below.

Definitions:

Nodes are discrete points along the path where states and controls are defined. The entire continuous path is represented by the states and controls at these discrete points. The states and controls at the all the node points form part of a vector making up the NLP variable that inputs to the NLP solver. Node locations are unevenly distributed and are pre-determined at the LGL points. LGL points occur at the zeros of the first derivative of the Legendre polynomials in order to minimize the mean-squared error in Lagrange interpolating polynomial approximation of states and controls.

Knots are double node points that are part of the optimization process. Trajectory end points are also called knots. Interior knots are used to define intermediate events where there may or may not be state discontinuities. LGL node distribution always concentrates more nodes next to knots. Intermediate or interior knots have a left and right side where state constraints are imposed.

Events are the set of conditions that define the manifolds which constrain the path (at the knots). The event manifolds usually consist of the boundary conditions; in this

thesis they represent the locations and velocities of Earth and Mars along their respective orbits in space and time. Within the DIDO structure, knots are the primary way to represent an event condition. An example of event boundary conditions is shown below where **e** represents the event vector, and $\mathbf{e}_l$ and $\mathbf{e}_u$ represent the lower and upper bounds of the vector respectively.

$$\mathbf{e}=\begin{bmatrix}\mathbf{r_0}\\ \mathbf{v_0}\\ \mathbf{r_f}\\ \mathbf{v_f}\\ \vdots\end{bmatrix} \text{ where } \mathbf{e}_l=\begin{bmatrix}\mathbf{r_{0}}_l\\ \mathbf{v_{0}}_l\\ \mathbf{r_{f}}_l\\ \mathbf{v_{f}}_l\\ \vdots\end{bmatrix} \text{ and } \mathbf{e}_u=\begin{bmatrix}\mathbf{r_{0}}_u\\ \mathbf{v_{0}}_u\\ \mathbf{r_{f}}_u\\ \mathbf{v_{f}}_u\\ \vdots\end{bmatrix}$$

such that

$$\mathbf{e}_l \le \mathbf{e}(\mathbf{x}(t_0),\mathbf{x}(t_i),\mathbf{x}(t_f),\mathbf{p},t_0,t_i\ t_f) \le \mathbf{e}_u$$

where the vector **p** contains any static parameters that may constrain the events. Events form the constraints that DIDO supplies to the generic NLP solver and are listed in tables for the various OCPs in this thesis.

Path constraints, **g**, are the state-control constraints that couple control and state variables in the following manner.

$$\mathbf{g}_l \le \mathbf{g}(\mathbf{x}(t),\mathbf{u}(t),t) \le \mathbf{g}_u$$

For numerical reasons it is sometimes advantageous to represent the solar sail controls as the sine and cosine components of the cone and clock control angles instead of the actual angles themselves. The path function in this case may be represented as $\mathbf{g}=\begin{bmatrix}u_1^2+u_2^2-1\\ u_3^2+u_4^2-1\end{bmatrix}$ where $\mathbf{g}_l=\mathbf{g}_u=\mathbf{0}$ forming an equality constraint. The actual control angles may be recovered by taking the arctangent of coupled controls (using Matlab's atan2 function serves well here).

Bounds are the upper and lower limits on states, controls or times. The resulting restriction is in the form of an equality or inequality constraint. For instance, a set of state variables may be bounded by $\mathbf{x}\in S\subset\mathbb{R}^N$ usually written in the form $\mathbf{x}\in[\mathbf{x}_{lb},\mathbf{x}_{ub}]$ or $\mathbf{x}_{lb}\le\mathbf{x}\le\mathbf{x}_{ub}$. Equality constraints are implemented simply by making $\mathbf{x}_{lb}=\mathbf{x}_{ub}$.

Within DIDO, states and controls are discretized, so variables are bounded at the node points.

Event Constraints are those linear or non-linear conditions that restrict states at the events. For instance, in an orbital rendezvous problem, the final spacecraft velocity is coupled with the final radius of the target planet according to the relation $v_f = \frac{m}{\sqrt{r_f}}$, forming a non-linear event constraint. When a state at a particular event condition is unconstrained, it is considered ***free.*** For example, when intercepting Mars in minimum time, we do not specify where in space or time the intercept must occur, thus final time and final angular displacement are free variables.

B. VALIDATING SOLUTIONS

Once a solution to a particular trajectory was obtained, it was important to see if the path was physically *feasible* and if so was it *optimal.* Checking for feasibility proved to be relatively straightforward, while verifying optimality was a bit thornier. This process linked the stark numerical optimization solution with a physical spacecraft trajectory that "made sense".

1. Feasibility

A glance at the state data at the knots would verify that the event constraints had been met. In order to verify that the *entire* state history conformed to physical laws governed by the equations of motion we use a propagator. Taking the control history generated by the numeric optimization and the initial conditions, the path of the sail could be propagated by means of a numeric ordinary differential equation solver on the non-linear equations of motion. The propagation used the same dynamics employed by the DIDO code to verify that DIDO was properly applying the dynamic constraints to the path. To verify the equations of motion in the dynamic model used by DIDO, the equations were propagated without controls to generate familiar Keplarian orbits. The state and control history figures in this thesis generally show the DIDO states represented

by markers at the discrete node points and the propagated solution as a line that will pass through DIDO's markers when the solutions are consistent.

2. Optimality

After testing for feasibility, the solution was checked for *optimality.* The primary method of verifying that an optimal trajectory had been achieved was observing the behavior of the Hamiltonian, H(**x**, **u**, λ, t). DIDO-derived Hamiltonian values were generated at the predetermined node points for simple problems (i.e. no intermediate knots) and were used to verify minimum time optimal controls. Since none of the problems posed in this thesis have a Hamiltonian as explicit function of time, then we can say

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} = 0$$

Hamiltonians that are constant with respect to time are necessary (but not sufficient) conditions for optimality in these cases. Moreover, the optimal control solution is the one where H is minimized with respect to controls, **u** thus defining an NLP problem with the Lagrangian of the Hamiltonian (for details see Appendix A). For minimum time problems, $H = -1$ for all time. The first and second order necessary (but not sufficient) conditions for optimality using the Hamiltonian are written as

$$\frac{\partial H}{\partial \mathbf{u}} = 0 \text{ and } \frac{\partial^2 H}{\partial \mathbf{u}^2} \geq 0$$

Additionally, dual variable outputs were used with the derived solar sail control law to produce controls that could be compared with the DIDO-derived controls. This would test if the optimal controls were conforming to the analytical optimal control steering law solution. Occasionally there existed one or more local minimums in which case varying the initial guess assisted in finding a better local minimum.

3. Comparison with Other Optimal Solutions

Often it serves well to verify DIDO-generated optimal solutions that use the pseudospectral method with solutions created using other methods. As mentioned before,

there are no known optimal control solutions to the actual solar sail cycler, however there are solutions to the simpler "benchmark" problems. One of the reasons for using the methodical build-up approach using simple building blocks in the form of benchmark problems was to verify that the coded models were producing verifiable results before proceeding to the uncharted waters. Comparisons are presented in the benchmark problems when other solutions are available.

THIS PAGE INTENTIONALLY LEFT BLANK

III. DEVELOPING THE OPTIMAL TWO-DIMENSIONAL CYCLER

A. TWO-DIMENSIONAL MODELS

1. Sail Model

A solar sail is a very large thin lightweight structure that transfers momentum from inbound solar photons to the spacecraft. Newton's laws dictate that the sail will accelerate twice as fast when photons are reflected than when they are absorbed. Thus, for maximum effectiveness the sail must be made of a very reflective material. For this analysis, the sail is modeled as a flat film (non-billowing sail) with a perfectly specular-reflecting surface on one side only. To characterize sail performance, we use the sail lightness number, b, defined as the ratio of the acceleration from solar radiation pressure (SRP) to the acceleration from the sun's gravity (ref 9 pp. 38-40).

$$b = \frac{a_{srp}}{a_{grav}} = \frac{\frac{2W_e R_e^2}{mc} A\cos^2 a}{m}$$

where W_e is the solar energy flux at 1 AU, R_e is the Sun-Earth distance, A is the sail area, a is the angle of the sail with respect to the Sun and m is the solar gravitational parameter[9]. Both accelerations are proportional to the inverse square of the radial distance from the sun, so b is an apt indicator of sail performance independent of location. A lightness number of b =.17 was used to model a reasonably high performance solar sail for most problems in this thesis for comparison to other solutions using the same sail[10]. Often the characteristic acceleration[11], a_0, is used to describe the performance of the sail instead the sail lightness number. This characteristic acceleration is the change in velocity that the sail would experience at a one AU distance from the sun when the sail exposes all of its area toward the sun. The relationship between the lightness number and the characteristic acceleration is

$$b\frac{m}{r_0^2} = a_0$$

For example, a sail lightness number of $b = .17$ corresponds to a characteristic acceleration of $a_0 = 1\frac{mm}{s^2}$.

The sail control angle, a, is defined as the angle between the sun-sail position vector, **r**, and the sail normal, **n** called the cone angle (Figure 4). This angle determines both the magnitude and direction of a solar force imparted to the spacecraft. Notice that when $a = 0$, maximum thrust is achieved and when $a = 90^o$, thrust is zero. The sail can never impart a force toward the sun (only gravity can do this). To ensure the sail will not flip over exposing its non-reflective side, the control angle is bounded by

(1) $$-\frac{p}{2} \le a \le \frac{p}{2}$$

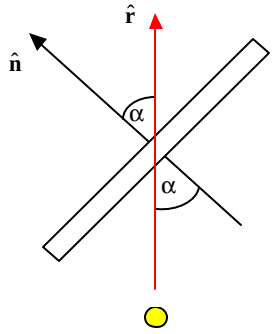

Figure 4 Sail normal, solar radial vector and control angle a**.**

2. Dynamic Model

Using the polar coordinates in the heliocentric frame, the two-dimensional equations of motion are derived in Appendix B. Given the state vector $\mathbf{x} = [r, q, u, v]^T$, the equations of motion are expressed as $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$ where control $\mathbf{u} = a$,

$$\mathbf{f} = \begin{bmatrix} u \\ \frac{v}{r} \\ \frac{v^2}{r} - \frac{m}{r^2} + \frac{bm}{r^2}\cos^3 a \\ -\frac{uv}{r} + \frac{bm}{r^2}\cos^2 a \sin a \end{bmatrix} \tag{2}$$

The states r, θ, u, and v are the radial distance, angular displacement, radial velocity and transverse velocity respectively. The angular displacement, θ, is decoupled from the other equations and may be excluded in special cases making $\mathbf{x} \in \mathbb{R}^3$. For numerical analysis of some problems, θ was retained as a state for proper intercept phasing during encounter events (more on that later).

Having thus imposed the dynamic constraints *between* the desired manifolds, we seek to constrain the conditions *at* the manifolds. This is accomplished by setting the constraints of the desired events.

3. Events Model

All the benchmark missions modeled here share the same cost function, dynamic equations of motion and sail control model. The distinguishing feature of each mission is its particular event function. The events are arranged into an event function vector **e** such that $\mathbf{e} \in \mathbb{R}^{Np}$ where N_p is the number of event constraints. Elements of **e** may be linear or nonlinear expressions and may be bounded as equality or inequality constraints. These event constraints for the flyby and rendezvous missions are summarized in the following table where a_m is the semi-major axis of Mars' orbit (1.524 AU).

Event	*Flyby [lb,ub]*	*Rendezvous [lb,ub]*
r_0	[1,1]	[1,1]
q_0	[0,0]	[0,0]
u_0	[0,0]	[0,0]
v_0	[1,1]	[1,1]
r_f	$[a_m, a_m]$	$[a_m, a_m]$
q_f	free	free
u_f	free	0
$v_f - \frac{1}{\sqrt{r_f}}$	free	[0,0]

Table 2 Event Conditions for Flyby and Rendezvous Problems

Notice that all the initial states (denoted by the '0' subscript) are equality constraints with the same lower and upper bounds (***[lb,ub])*** indicating that the spacecraft will launch from a circular Earth orbit at Earth's orbital speed. The final rendezvous along-track velocity (denoted with the "f" subscript) is the only non-linear constraint. These events are assembled within DIDO into a constraint vector, $\mathbf{e} = [r_0\ q_0\ u_0\ v_0\ r_f\ q_f\ u_f\ v_f{}^2 r_f - 1]^T$. The last element is written in a different form to avoid the +/- ambiguity using the square root.

B. THE OPTIMAL CONTROL PROBLEM

We now come to posing the solar sail optimal control problem (OCP). This is where one must question what exactly are we trying to do. With most conventional propulsion schemes, it is desirable to minimize fuel consumption. Solar sails, however, require no propellant so the goal of most of the solar sail trajectories presented here is to accomplish a given mission in minimum time. The general OCP is stated as follows[3].

Minimize the Bolza cost functional:

$$J[\mathbf{x}(\cdot),\mathbf{u}(\cdot),t_0,t_f;\mathbf{p}] = E(\mathbf{x}(t_0),\mathbf{x}(t_f),t_0,t_f;\mathbf{p}) + \int_{t_0}^{t_f} F(\mathbf{x}(t),\mathbf{u}(t),t;\mathbf{p})dt$$

Subject to:

Dynamic constraints $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t),\mathbf{u}(t))$

where $\mathbf{f}$ is given in equation (2) and $\mathbf{u}$ are the controls,

Path constraints $\mathbf{g}_l \le \mathbf{g}(\mathbf{x}(t),\mathbf{u}(t),t) \le \mathbf{g}_u$

Event constraints $\mathbf{e}_l \le \mathbf{e}(\mathbf{x}_0, \mathbf{x}_i, \mathbf{x}_f, t_0, t_i, t_f; \mathbf{p}) \le \mathbf{e}_u$

and bounds on state and control variables

$$\mathbf{x}_l \le \mathbf{x}(t) \le \mathbf{x}_u$$
$$\mathbf{u}_l \le \mathbf{u}(t) \le \mathbf{u}_u$$

With an established coordinate system and scaling system, we are postured to model the sail, dynamics and events. These models are structured into a format ready for DIDO to use as the OCP constraints. Before attempting a numerical solution, we desire the analytical solar sail steering law to help verify the solution as described below.

C. SOLAR SAIL CONTROL LAW

Following the guidelines set forth by the minimum principle (Appendix A), we start by establishing the cost function. To minimize time, we choose to express the cost functional from the previous section in Lagrange form.

$$E = 0, F = 1$$

The state space vector and its derivative are written as

$$\mathbf{x} = \begin{bmatrix} r \\ q \\ u \\ v \end{bmatrix} \quad \text{and} \quad \mathbf{f}(\mathbf{x},\mathbf{u}) = \begin{bmatrix} u \\ \frac{v}{r} \\ \frac{v^2}{r} - \frac{m}{r^2} + \frac{bm}{r^2}\cos^3 a \\ -\frac{uv}{r} + \frac{bm}{r^2}\cos^2 a \sin a \end{bmatrix}$$

where the control is given by the sail cone angle $\alpha(t)$.

Notice that θ is uncoupled from the other equations, therefore it will be ignored for the rest of the sail control analysis. Thus, we write the Hamiltonian as

$$H(\mathbf{x},\mathbf{u},\lambda) = F + \lambda^T \mathbf{f} = 1 + \lambda_r u + \lambda_u \left[\frac{v^2}{r} - \frac{1}{r^2} + \frac{b}{r^2}\cos^3\alpha \right] + \lambda_v \left(-\frac{uv}{r} + \frac{b}{r^2}\cos^2\alpha \sin\alpha \right)$$

where λ is the costate vector and the solar gravitational parameter has been normalized to unity.

The Lagrangian of the Hamiltonian is therefore

$$L(\mathbf{x},\mathbf{u},\lambda) = H + \mu g = 1 + \lambda_r u + \lambda_u \left[\frac{v^2}{r} - \frac{1}{r^2} + \frac{b}{r^2}\cos^3\alpha \right] + \lambda_v \left(-\frac{uv}{r} + \frac{b}{r^2}\cos^2\alpha \sin\alpha \right) + \mu_\alpha \alpha$$

where μ_α is now the covector associated with the path constraint $-\frac{\pi}{2} \le \alpha \le \frac{\pi}{2}$..

Applying the minimum principle yields

$$\frac{\partial L}{\partial \alpha} = -\lambda_u \frac{3b}{r^2}\cos^2\alpha \sin\alpha + \lambda_v \left[-\frac{2b}{r^2}\cos\alpha \sin^2\alpha + \frac{b}{r^2}\cos^3\alpha \right] + \mu_\alpha = 0$$

so μ_α may be written as

$$\mu_\alpha = \frac{b}{r^2}\cos\alpha \left(3\lambda_u \cos\alpha \sin\alpha + 2\lambda_v \sin^2\alpha - \lambda_v \cos^2\alpha \right)$$

For the given control constraints, the KKT conditions give

$$\alpha = -\frac{\pi}{2}, \mu_\alpha \le 0$$

$$\alpha = +\frac{\pi}{2}, \mu_\alpha \ge 0$$

$$\frac{-\pi}{2} < \alpha < \frac{\pi}{2}, \mu_\alpha = 0$$

Limiting the analysis to the interior controls where $\frac{-\pi}{2} < \alpha < \frac{\pi}{2}$ and μ_α=0, we can obtain the control law.

$$\left(3l_u \cos\alpha \sin\alpha + 2l_v \sin^2\alpha - l_v \cos^2\alpha\right) = 0$$

Dividing the equation by $\cos^2\alpha$ produces a quadratic equation with $\tan\alpha$ as the independent variable.

$$2l_v \tan^2\alpha + 3l_u \tan\alpha - l_v = 0$$

(3)
$$\boxed{\tan\alpha = \frac{-3l_u \pm \sqrt{9l_u^2 + 8l_v^2}}{4l_v}}$$

This tangent control law will not have quadrant ambiguity since we limited the control angle by $-\frac{\pi}{2} \le \alpha \le \frac{\pi}{2}$.

D. RESULTS

Before analyzing results, it is useful to correlate certain key sail angles with physical meaning. As mentioned in the discussion of the sail model, the thrust magnitude due to the solar radiation pressure (SRP) on the sail is not a free control variable as it is with many conventional propulsion systems. The acceleration imparted by the sail, T_s/m, has a magnitude that is dependent on radial distance and the control angle, α, according to the relationship $\frac{T_s}{m} = \frac{\beta\mu}{r^2}\cos^2\alpha$. Notice that for a given r, maximum radial thrust occurs when $\alpha = 0$ and it's sail area is totally exposed to the sun. Alternatively, the sail is effectively "off" and the trajectory becomes ballistic when $\alpha = \pm\pi/2$. The sail can never isolate tangential thrust from radial thrust. Radial thrust is present for all control angles except for $\alpha = \pm\pi/2$ when the sail edge is toward the sun. To assist in interpreting the control profile, we determine the control angle at which maximum transverse acceleration occurs. The transverse acceleration is given by:

$$a_t = \frac{T_s}{m}\sin\alpha = \frac{\beta\mu}{r^2}\cos^2\alpha \sin\alpha, \quad -\frac{\pi}{2} < \alpha < \frac{\pi}{2}$$

Setting the derivative with respect to α to zero *for a given r* yields the control angle providing maximum transverse acceleration.

$$\frac{da_t}{da} = -2\cos a \sin^2 a + \cos^3 a = 0 \qquad (4)$$
$$a = .615 = 35.2°$$

1. Benchmark Problem Solutions with *Circular* Coplanar Orbits

a. Earth-Mars Flyby

The purpose of solving the flyby mission serves several purposes. First, the solution may be compared to other known solutions to verify that the numerical analysis using DIDO works properly. Second, the minimum time of flight serves as a lower time bound for future coplanar trajectory optimization problems. Third, groundwork is established to model a bounded initial C3 from Earth allowing a limited "boost" from a conventional rocket at time t = 0.

Given an initial guess of the initial and final positions and velocities, DIDO outputs the minimum time Mars flyby states and controls (plotted as markers in Figure 5). For a minimum time flyby path between circular coplanar orbits, a sail with lightness number β=.17 takes 0.45 years. The probe sails past Mars at a speedy 8.7 km/s (relative to Mars). Sail attitude favors a large local transverse acceleration profile initially to rapidly build radial acceleration, and then gradually exposes more sail area to the sun throughout the maneuver until the sail normal is parallel with the sun's rays (Figure 6). DIDO costate outputs representing $\lambda = [\lambda_r, \lambda_q, \lambda_u, \lambda_v]^T$ are plotted in Figure 7, and may be used with the tangent steering control law in equation (3) to generate derived controls. The true anomaly costate, λ_q, is zero as expected since q does not appear in the Hamiltonian. A comparison of costate-derived controls with the DIDO output controls appears in Figure 8.

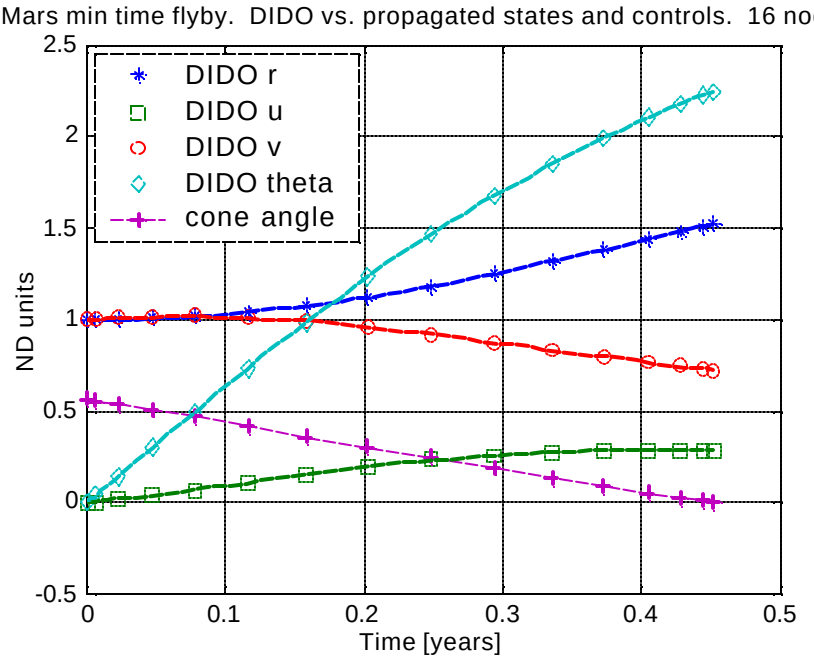

Figure 5 States and Control for Mars Flyby Mission. Markers are DIDO output and lines represent propagated path.

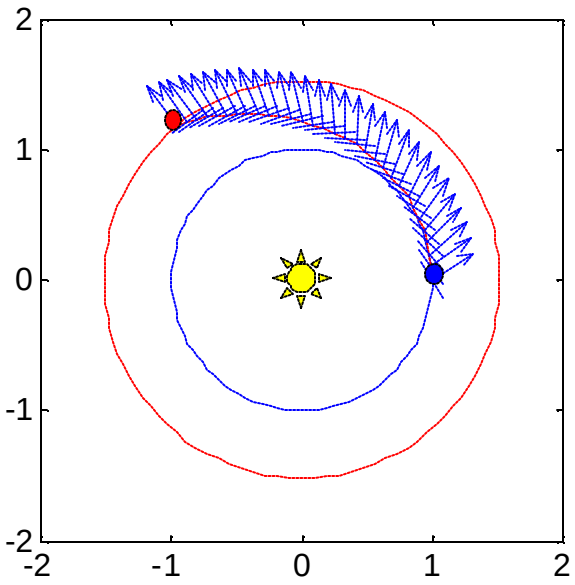

Figure 6 Mars Flyby Trajectory with Sail Profile and Normal.

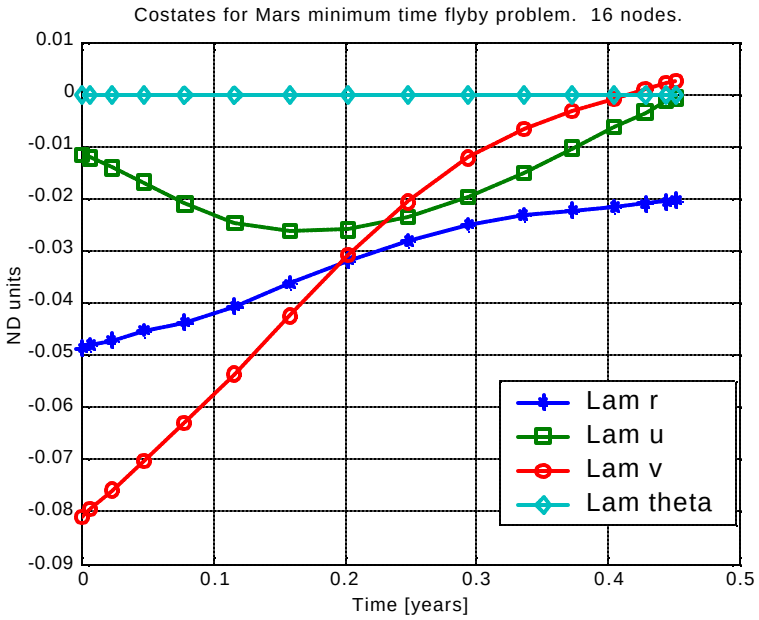

Figure 7 Mars Flyby Mission Costates

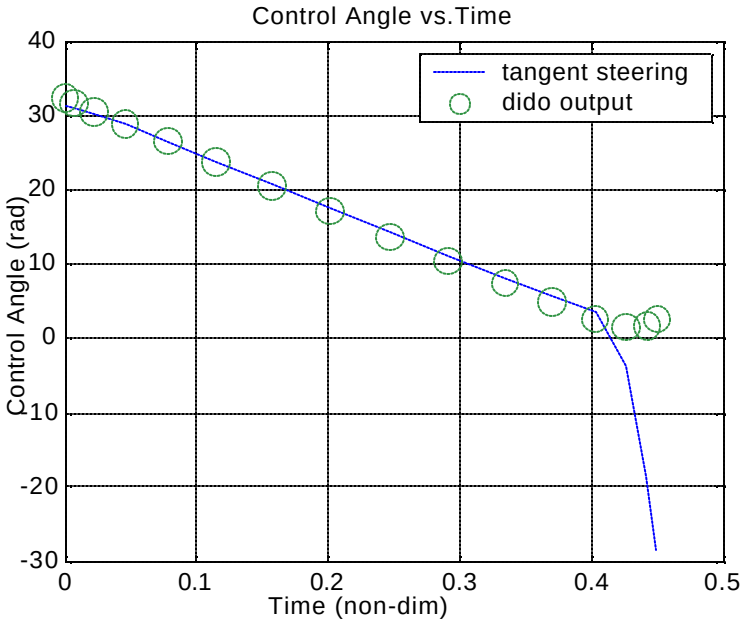

Figure 8 Mars Flyby Cone Angle Control

Thus far, the flyby model has restricted the sail to start under its own power without the aid of a kick motor. Modeling a kick motor is accomplished by changing the initial conditions in Table 2 for the u and v velocity components to be constrained variables according to the following relationship.

$$0 \le \left|\mathbf{V}_0\right| \le \Delta V_{\max}$$

where $\mathbf{V}_0 = [u_0, v_0]^T - \mathbf{V}_e$. Not surprisingly, the optimal solution makes use of whatever $\Delta V_{\max}$ is permitted to intercept Mars in the quickest time, although the direction of departure depends on the maximum allowable kick. The mission is accomplished much faster than the previous sail-only solution. For example, when $\Delta V_{\max} = 6\frac{km}{s}$ the optimal solution uses all the initial help it can get reducing the time to intercept to only .22 years. We find, however, that the same behavior does not apply to the rendezvou s problem with its final velocity constraint.

Feasibility of the solution is demonstrated in Figure 9 depicting the path propagated using an ODE solver given the initial conditions and the DIDO -derived

control profile. Although difficult to rigorously *prove* that this is the optimal trajectory, we can at least show that necessary conditions for optimality are satisfied in accordance with the discussion in the section on Optimality. The Hamiltonian is constant with respect to time (Figure 10) and the second order condition $\frac{\partial^2 H}{\partial a^2} \geq 0$ is true (Figure 11).

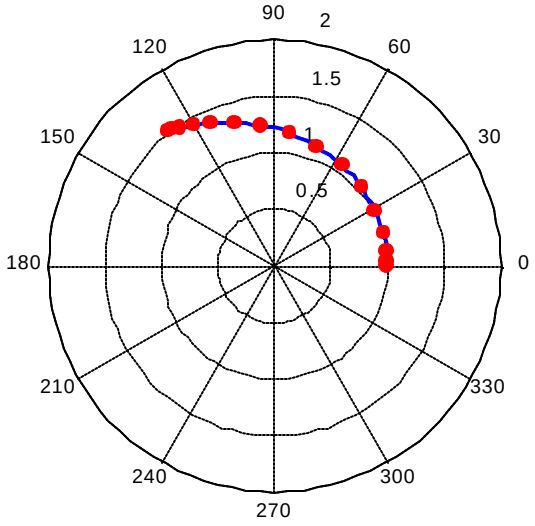

Figure 9 Mars Flyby Mission Propagation (line) with DIDO Output (dots)

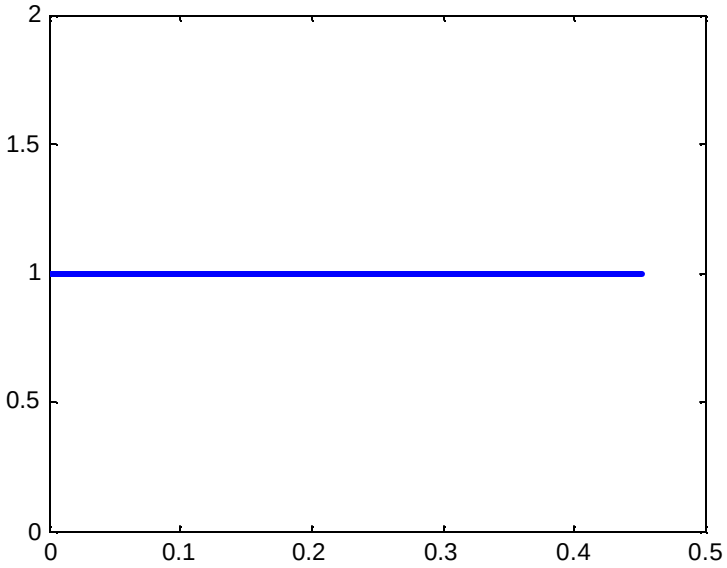

Figure 10 Hamiltonian Profile for Mars Flyby Mission

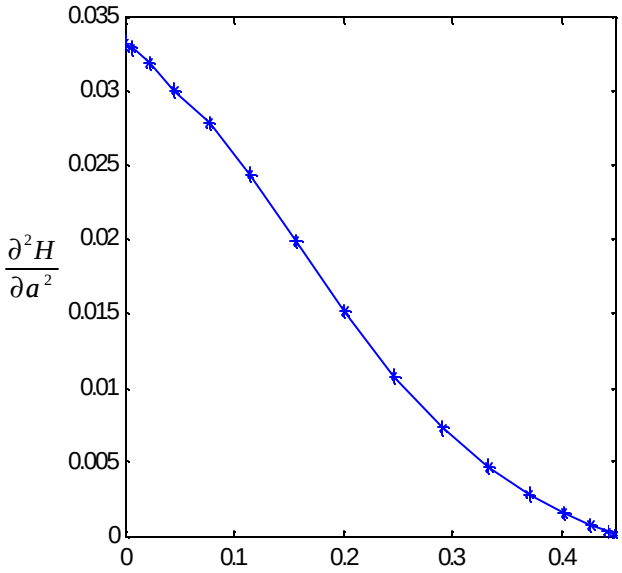

Figure 11 Second Order Condition for Mars Flyby

b. Rendezvous

As expected, the Mars rendezvous mission takes longer to accomplish than the flyby mission, requiring a full 1.11 years, which is consistent with the solution presented in ref 12, (Figure 12). The spacecraft final velocity is constrained to match Mars' orbital velocity, which costs the sail extra transit time to accomplish. A surprising feature of the trajectory is that the probe sweeps an arc outside of the circular orbit of Mars (Figure 13). This maneuver is observed in more complex minimum time solar sail missions as well since it appears to make the best use of radial and transverse thrust. There is a point on the outbound path (almost a quarter of the way into the mission time) when the control angle rotates from a small angle, where much of the sail area is exposed to the sun, to a more edge-on aspect. The sail attitude gradually rotates after this to favor more transverse acceleration as it sweeps past Mars' orbit radius and approaches Mars. By the time the sail has reached Mars, the control angle is 5° greater than its maximum transverse velocity setting.

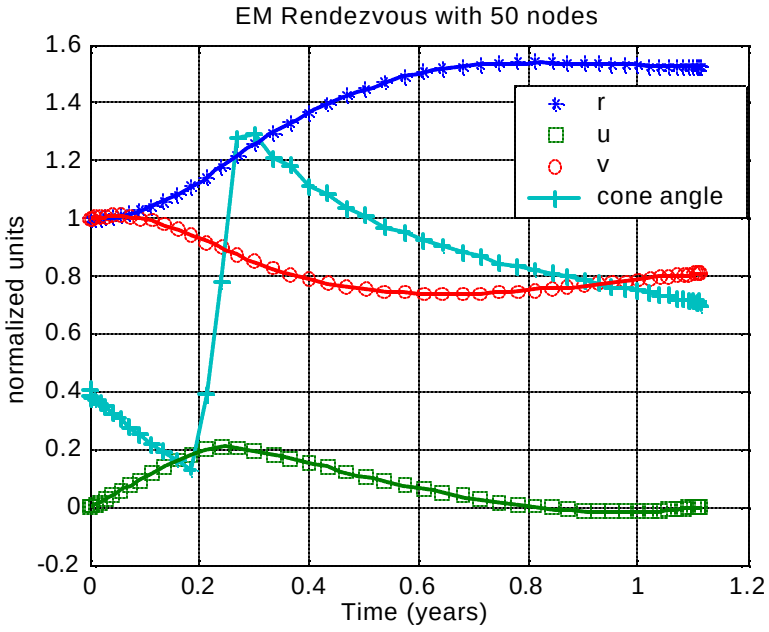

Figure 12 DIDO State and Control Output (Markers) and Propagated Path (line) for EM rendezvous.

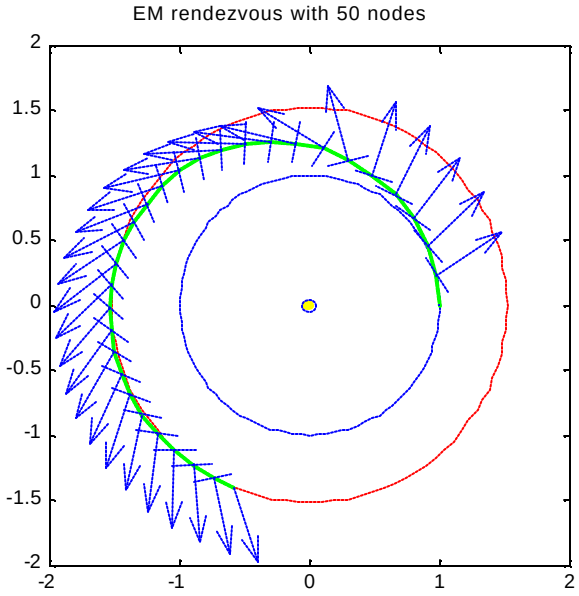

Figure 13 Sail Trajectory and Profile for Earth-Mars Rendezvous

Interpreting the costate history provides a more complete physical analysis of the optimal trajectory. Each costate in Figure 14 represents a Lagrange multiplier that signifies instantaneous sensitivities of the cost function to instantaneous variations in the corresponding state. Once again, the costate corresponding to θ is zero since this state is completely uncoupled from the other states in the dynamic equations and never appears in the Hamiltonian. Several observations may be made regarding the critical point in the first quarter of the trajectory when the control angle makes a radical change. Recalling the optimal sail steering law in equation (3) it is seen that when $l_u = 0$, then the maximum transverse velocity profile occurs. This value would occur when $\tan a = \frac{\sqrt{2}}{2}$ ($a = 35.2^\circ$). Also, when $l_v \to 0$ then $a \to 90^\circ$. Observing the costates for the Earth-Mars (EM) rendezvous, we note that l_u crosses zero at t = 1.225 years and soon afterwards l_v reaches its closest point to zero. This is consistent with a rapid change of sail attitude at that time during the mission and corresponds to when the sail reaches maximum radial velocity.

Lagrange multipliers along with state outputs are fed into the derived tangent steering law (equation (3)) to obtain a control profile. The resulting control profile is shown in Figure 15 and is plotted with the DIDO control output for comparison.

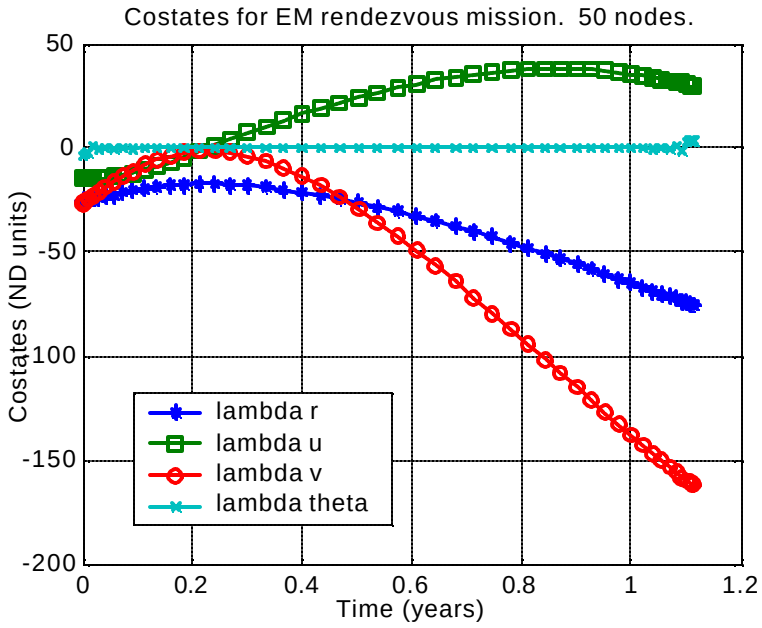

Figure 14 Costates for Earth-Mars Rendezvous Mission

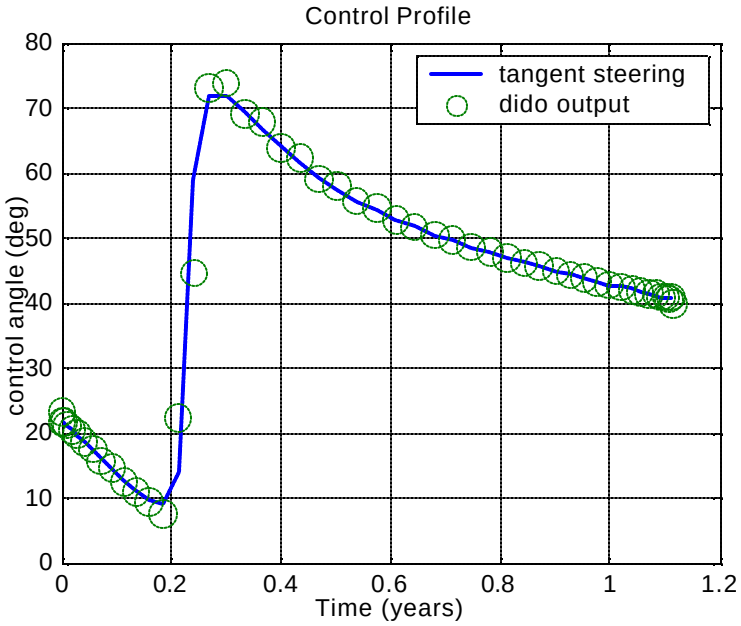

Figure 15 Comparison of DIDO Controls and Tangent Steering Control Law for EM Rendezvous.

Propagating the spacecraft with the given control history produces the path shown in Figure 17 with DIDO states at node points shown for comparison. Testing the second order necessary conditions for optimality we observe in Figure 16 that $\frac{\partial^2 H}{\partial a^2} \geq 0$ for all time and also the Hamiltonian is fairly constant (Figure 18).

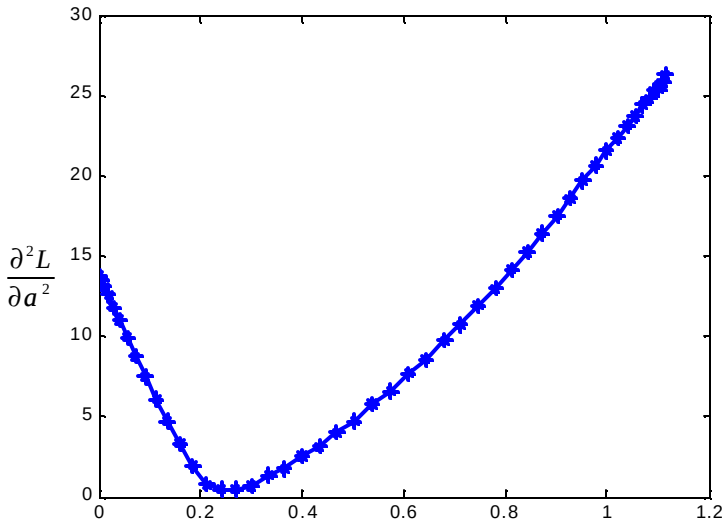

Figure 16 Second Order Necessary Condition $\frac{\partial^2 H}{\partial a^2} \geq 0$

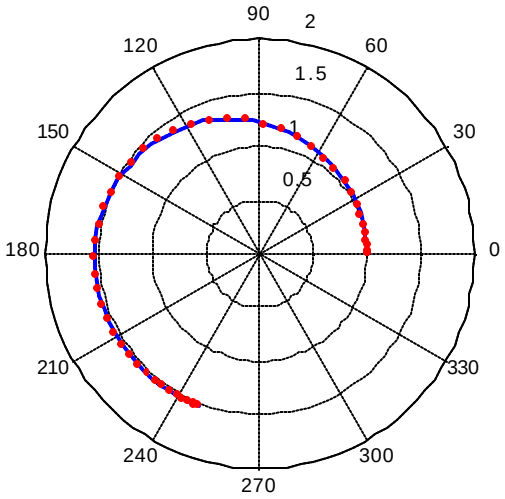

Figure 17 Propagated Path (line) with DIDO State Output (dots) for EM Rendezvous

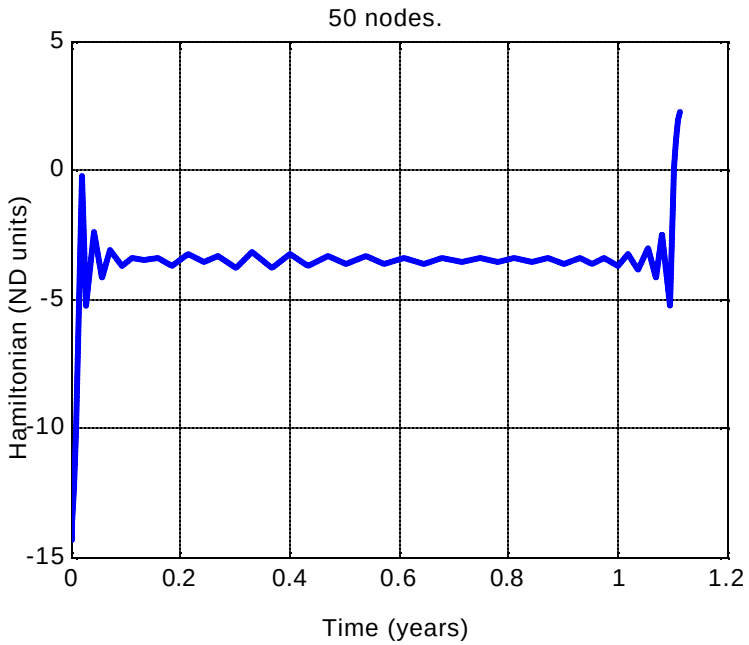

Figure 18 Hamiltonian for Minimum Time EM rendezvous.

What happens when we reverse the problem such that we seek the minimum time from Mars ($V_{\infty}|_{mars} = 0$) to Earth rendezvous? It turns out that swapping initial conditions with final conditions generates a time-reversed state history and a control profile that is inverted and reversed. The experiment may be carried out within the DIDO structure by simply reversing the event conditions.

Event	*ME Rendezvous [lb,ub]*
r_0	$[a_m, a_m]$
q_0	free
u_0	[0,0]
v_0	$\left[\frac{1}{\sqrt{r_0}}, \frac{1}{\sqrt{r_0}}\right]$
r_f	[1,1]
q_f	free
u_f	[0,0]
v_f	[1,1]

Table 3 Mars-Earth Rendezvous Event Conditions.

The state history of the optimal Mars-Earth rendezvous is a mirror image of the Earth-Mars rendezvous (Figure 19 and Figure 20). Time of flight, Δt, and the change in angular displacement, $\Delta\theta$, are precisely the same for both profiles. Furthermore, the control profile is inverted and time reversed. This result leads to an even more interesting question; what happens when we desire the sail to launch from Earth, rendezvous with Mars, then immediately start its trek back home to rendezvous with Earth? This benchmark mission is the subject of the next section.

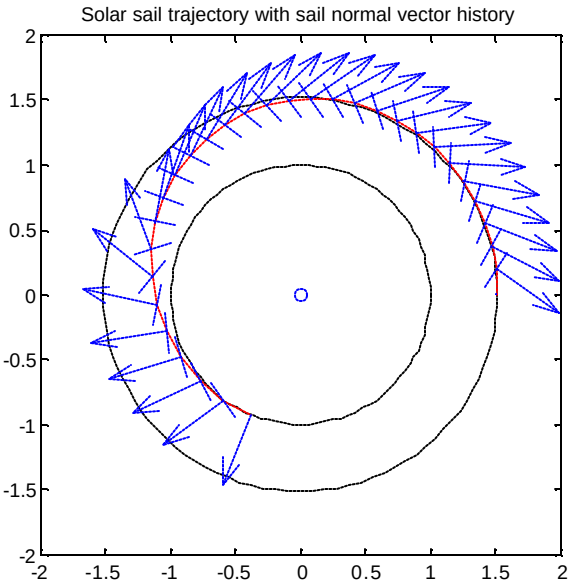

Figure 19 Mars to Earth Rendezvous. Shows reverse trajectory of EM rendezvous mission

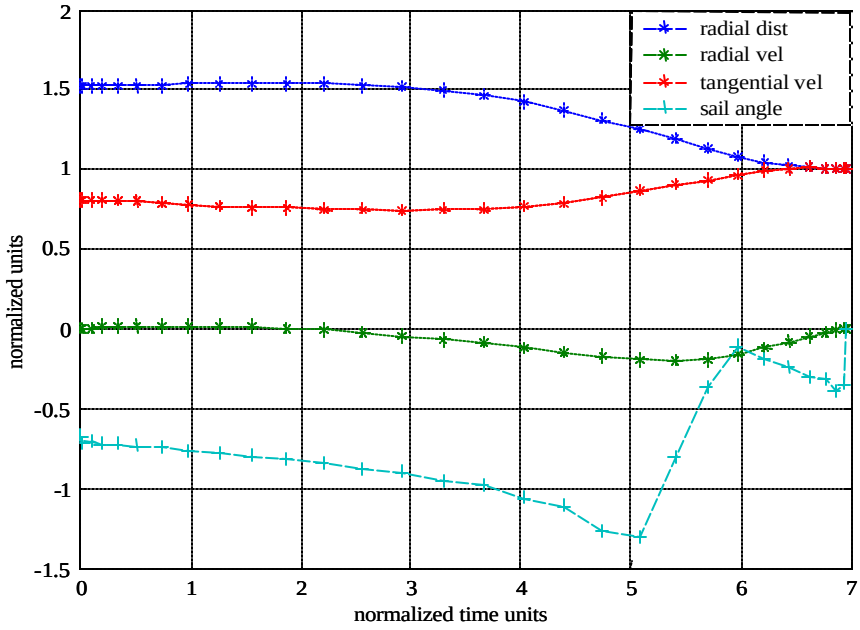

Figure 20 ME Rendezvous States and Controls

c. Double Rendezvous

As a next step toward modeling a cycler orbit, we explore the double rendezvous. In this mission we seek the minimum time that it takes to make an Earth-Mars-Earth (EME) round trip with a zero relative velocity at both planets for a given sail lightness number β. To study the behavior of the optimal trajectory, one assumes that Earth or Mars gravity will not significantly influence the craft. Low relative velocity is desirable while encountering the target planets since it requires less energy for a greeting taxi craft to intercept and dock with the passing spacecraft.

For the preliminary analysis, the initial, intermediate and final orbits are modeled as circular and coplanar. Earth phasing for the final Earth rendezvous event is accomplished by coupling the sail and Earth final positions using the Earth's mean motion and the time of flight.

The solution to the minimum time EM rendezvous problem provides a lower bound for the time that it must take to reach Mars during the double rendezvous problem. Since the fastest possible time to rendezvous with Mars with the given solar sail (β=.17) was determined to be 1.11 years from the previous section, this serves as the minimum time to encounter Mars halfway through the double rendezvous mission. Remembering that the reverse Mars-Earth (ME) rendezvous took exactly the same amount of time, we now have bounded the time to complete the whole EME double rendezvous. Setting the intermediate destination at the Mars orbit and the final destination with Earth and constraining the event velocities to match those of the respective planets we are in a position to solve the optimal EME double rendezvous. Shown in Table 4 is a summary of the event constraints where superscripts indicate before (-) or after (+) the intermediate knot representing the Mars encounter. There is a single non-linear event constraint that is responsible to ensure proper phasing (i.e. Earth makes more orbits around the sun than the sail does, but still meets with it in the end). The output states and controls are shown Figure 21 with the path shown in Figure 22. Essentially it turns out to be the EM minimum time rendezvous solution patched together with the ME rendezvous solution with slightly longer transit times. The most obvious features of the profile are the state symmetries and control antisymmetry about the mid-

maneuver point. Event conditions at the interior knot required only that the states be continuous and equal to Mars' state.

Event	*[lb,ub]*	*Event*	*[lb,ub]*	*Event*	*[lb,ub]*	*Event*	*[lb,ub]*
r_0	[1,1]	r_i^-	$[a_m, a_m]$	r_i^+	$[r_i^-, r_i^-]$	r_f	[1,1]
q_0	free	q_i^-	free	q_i^+	$[q_i^-, q_i^-]$	q_f	free
u_0	[0,0]	u_i^-	[0,0]	u_i^+	$[u_i^-, u_i^-]$	u_f	[0,0]
v_0	[1,1]	v_i^-	$[v_m, v_m]$	v_i^+	$[v_i^-, v_i^-]$	v_f	[1,1]
Non-linear constraint			$\cos(q_{e0} + n_e(t_f - t_0) - q_f)$				[1,1]

Table 4 Event Conditions for EME Double Rendezvous

For the given sail, the total time to complete the trajectory is 2.41 years with both legs of the journey taking 1.205 years. The reason that the EME problem takes longer than the patched EM-ME problem is that the final Earth rendezvous event must be phased with Earth's position at the final time. The individual EM and ME rendezvous solutions only served to provide a lower time bound for the EME problem, so imposing the constraint that the sail final position is phased with Earth is about 10% more than the lower bound.

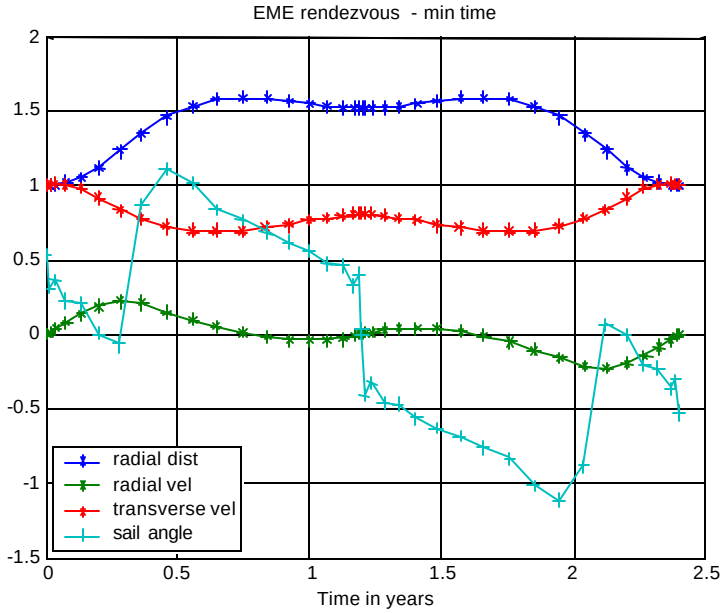

Figure 21 States and Controls for EME Double Rendezvous in Minimum Time (b=.17)

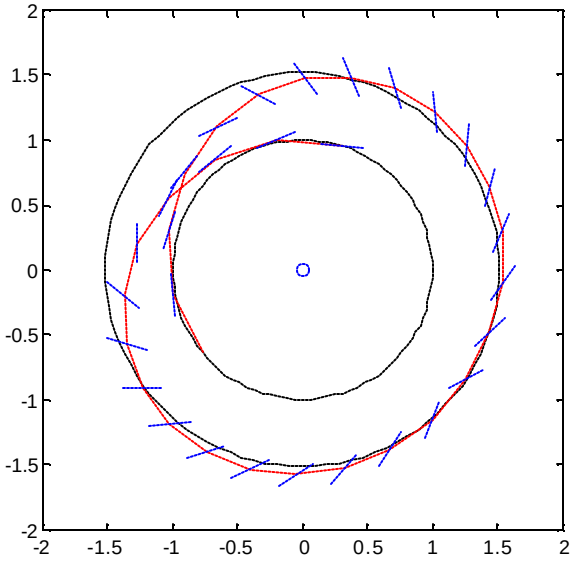

Figure 22 EME Double Rendezvous Trajectory and Sail Profile (b=.17)

This round trip solution only takes into account a quick trip out to graze the Mars orbit and then a return trip home. It may be desirable to have the probe linger with Mars for a span of time allowing for some sort of mission while traveling with the planet then returning back to Earth. The next question one may ask is the following. Given a sail of certain performance and Mars stay time, what is the minimum *transit* time EME trajectory (not including the Mars stay time)? For a given Mars stay time, some of the transit time to and from Mars is expended in performing a phasing maneuver to meet Earth at the final time. This characteristic phasing maneuver is exhibited in all double rendezvous problems with varying stay times but one – the one where the *stay time* spent with Mars happens to be just the right time to allow a return trip to Earth with no phasing maneuver required at all. As mentioned, this minimum possible return trip time is 1.1 years for sail lightness number β=.17. The following analysis shows that given the optimal time and angular displacement required for the EM rendezvous (and thus the ME rendezvous), we can deduce the Mars stay time that minimizes transit time. The optimal transit time and true anomaly traversed by the sail for an EM rendezvous (and ME rendezvous) are given by the following.

$$t^* = 1.116 \text{ years} \qquad q^* = 4.337 \text{ radians}$$

The following equations capture the motion of the Earth relative to that of the cycler craft where t_m = Mars stay time, Δq_e = change in Earth angular position and n_e and n_m represent Earth and Mars mean motion respectively.

$$\Delta q_e = n_e(2t^* + t_m) \tag{5}$$

Equation (5) indicates that Earth changes position by Δq_e during the same time that the sail transits outbound in t^* years, stays with Mars for t_m years, then returns home in t^* years. Recognizing the $n_e = 1\frac{rad}{TU}$ we can rewrite equation (5) as

$$\Delta q_e - t_m = 2t^* \tag{6}$$

During the round trip mission time the spacecraft traverses q^* radians on the outbound leg, $n_m t_m$ radians along with Mars, and q^* on the inbound leg.

(7) $$\Delta q = 2q^* + n_m t_m$$

All the while, Earth traverses the same span plus an extra N revolutions.

(8) $$\Delta q_e = \Delta q + 2p\, N_E \qquad (N_E = [0,1,2...])$$

Constraining the solution to include only the first Earth pass on the return (N=1), we substitute equation (7) into equation (8) and rearrange to obtain

(9) $$\Delta q_e - n_m t_m = 2q^* + 2p$$

Recognizing that the scaled mean motions of Earth and Mars are given by $n_e = 1$ and $n_m = \sqrt{\frac{1}{a_m^{\,3}}} = .534$ respectively, we may solve for Δq_e and t_m simultaneously using equations (6) and (9). This results in a Mars stay time of $t_m = 7.0114$ TUs = 117 days and an Earth span of $\Delta q_e = 16.03$ radians corresponding to 2.55 years of total mission time.

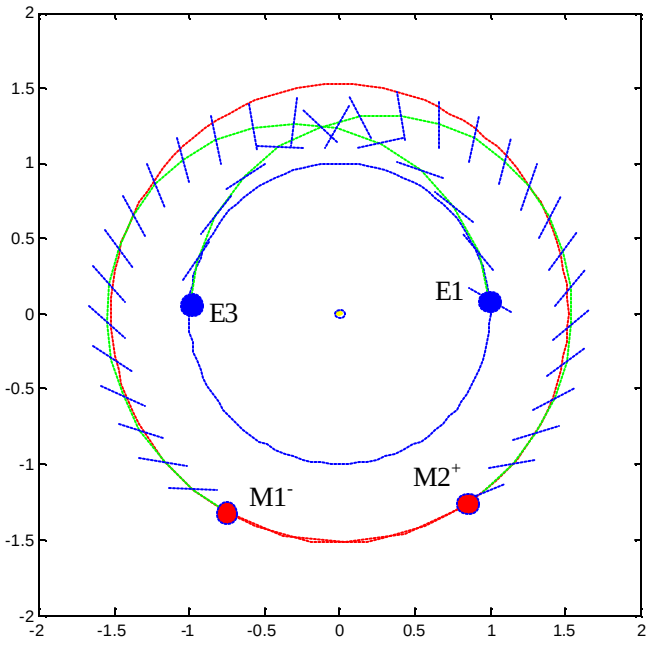

Figure 23 EME Double Rendezvous with Stay Time at Mars

When using the outbound and inbound transit times as a cost function, DIDO produces the output shown in Figure 23. The path mimics a patched minimum time EM-ME solution where the optimal Mars stay time is 123 days for the given sail (5% difference from the estimated value of 117 days). Any more or less time spent at the planet will either require a time-wasting phasing maneuver to allow Earth to catch up for a rendezvous, or a maneuver to catch up to the speeding target Earth. A plot showing the impact of various Mars stay times on mission transit times appears in Figure 24. Notice that it is safer to design a shorter-than-optimal Mars stay time into the mission. In the event of a schedule slide, the return transit time looks increasingly grim beyond the optimal point because the sail has a lot of space to cover as it attempts to catch up to Earth.

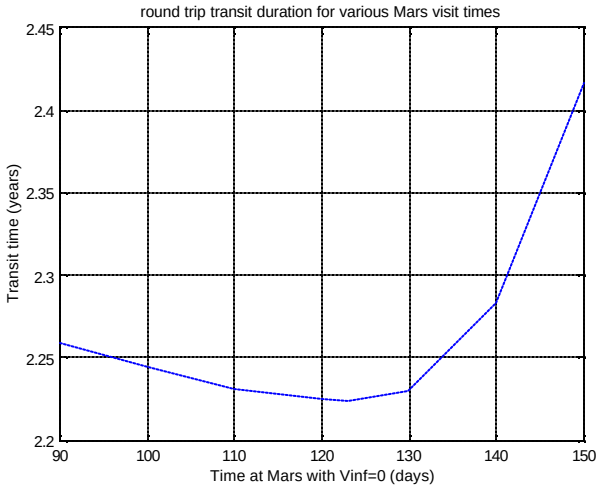

Figure 24 EME Double Rendezvous Transit times for Various Mars Stay Times. (b =.17)

The analysis thus far has only dealt with sails of lightness number β=.17. An interesting question that may be asked now is the following. Given a desir ed time of round trip flight with no stay time at Mars, what is the minimum size solar sail required? This question will be addressed later.

2. Benchmark Problem Solutions with an *Elliptic* Coplanar Orbit

Running the same battery of problems from the previous section using a higher fidelity model provides enormous insights into the characteristics of an optimal trajectory. Modeling trajectories between circular Earth and coplanar elliptic Mars orbits generates results that may be compared with the circula r orbits model to reveal the "knees in the curve" that the optimization process seeks in an effort to reduce the cost function just a little more. Knowing these characteristics of optimal solar sail trajectories will assist in understanding the behavior of optimal solar sail cyclers.

Because the final target manifold represents Mars' elliptical orbit, the new events model must relate the final Mars radial position, r_{mf}, and velocity, $\mathbf{V}_m$, with the final sail state, $\mathbf{x}_f$.

Constraining the initial and final sail angular positions to be coincident with the initial Earth and final Mars angular positions respectively, we obtain the following relationship for coplanar orbits.

$$q_{mf} = q_f - (w_m + \Omega_m)$$

where q_f is the final sail position, w_m is Mars' argument of periapsis and Ω_m is the right ascension of the ascending nodes.

In a perifocal system, Mars final polar coordinates r_{mf} and q_{mf} are related by

$$r_{mf} = \frac{p_m}{1 + e_m \cos q_{mf}} \tag{10}$$

where p_m in the numerator is the "parameter" or semi-latus rectum of the Mars orbit. The final event condition for radial position of the sail is simply

$$r_f = r_{mf} \tag{11}$$

For the rendezvous mission, we need to target the final Mars velocity as well. The speed of Mars at the final time is expressed as

$$V_{mf} = \sqrt{m\left(\frac{2}{r_{mf}} - \frac{1}{a_m}\right)}$$

where a_m is the semi-major axis of Mars. Components of the planetary velocity vector are defined in terms of the flight path angle b_m by

$$\cos b_{mf} = \frac{h_m}{r_{mf} V_{mf}} \text{ and } \sin b_{mf} = \frac{m}{h_m V_{mf}} e_m \sin\left(q_{mf}\right)$$

where h_m represents the magnitude of the Mars angular momentum vector.

The velocity vector resolved in the local vertical local horizontal frame (LVLH) produces

$$u_{mf} = V_{mf} \sin b_{mf}$$

$$v_{mf} = V_{mf} \cos b_{mf}$$

Representing final Mars and sail velocities by $\mathbf{V}_{mf} = [u_{mf}, v_{mf}]^T$ and $\mathbf{V}_f = [u_f, v_f]^T$ respectively, we obtain an additional constraint for the elliptic rendezvous mission

(12) $$\mathbf{V}_f = \mathbf{V}_{mf}$$

It now remains to run the same battery of battery of benchmark problems with the elliptic Mars event model.

a. Earth-Mars Flyby

Modeling Mars' orbit as an ellipse using eccentricity $e_m = .0935$ provides target radial distances that vary with the planets true anomaly (equation (10)). Earth's orbit eccentricity is only .0167, so the circular Earth orbit assumption is a good one (elliptic Earth orbit is considered in the 3-D model in the next chapter). Not surprisingly the time optimal mission selects a path that drives the sail from Earth orbit to Mars periapsis at high speed (Figure 25). This optimal path exploits Mars' eccentric orbit and presents the shortest distance from Earth to Mars, and thus the shortest transit time required since the final velocity is unconstrained. The time to intercept Mars using the standard sail ($b = .17$) is only 137 days, about 27 days faster than the corresponding circular orbits model (Figure 26).

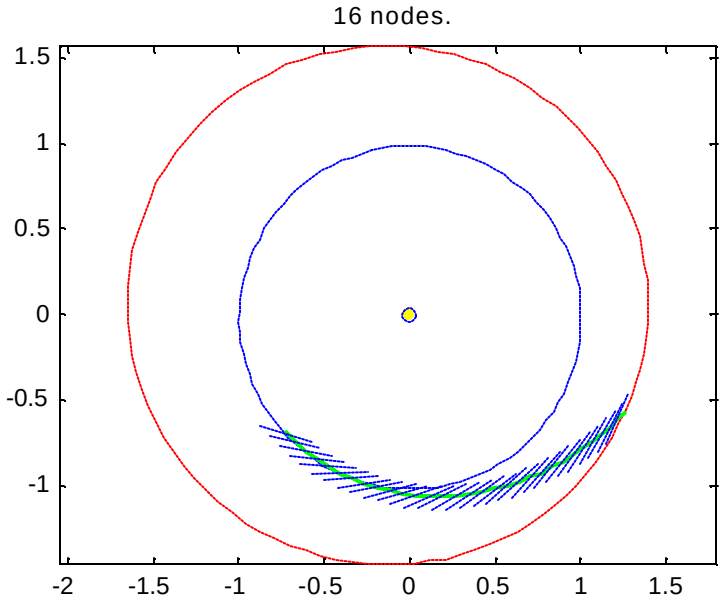

Figure 25 Minimum Time EM flyby with Mars Elliptical Orbit. Final position is at Mars periapsis.

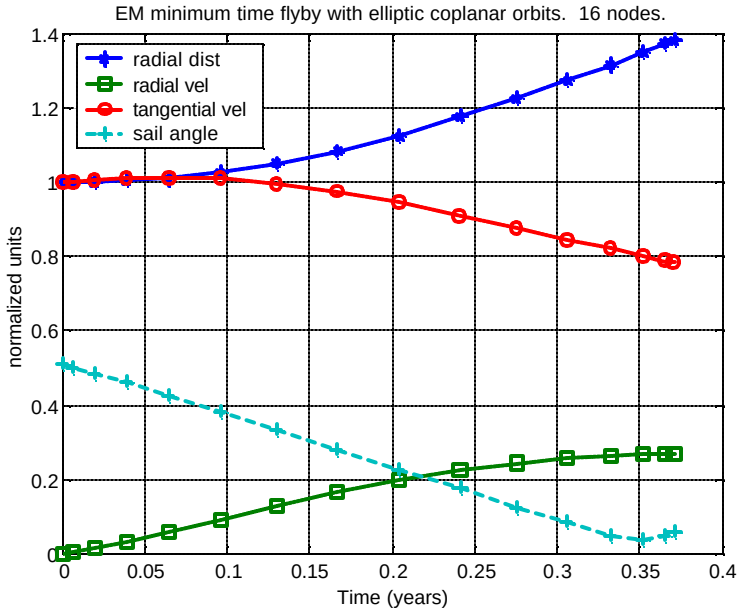

Figure 26 EM Flyby with Mars Elliptic Orbit

b. Rendezvous

To achieve the rendezvous mission in minimum time the path must account for target velocity as well as distance, both of which vary with respect to target planet true anomaly. Initial and final true anomalies, thus launch windows and arrival dates, are optimization parameters that seek to reduce the cost function by finding the optimal points on the target manifolds to bound the path. It is fascinating to observe that the optimal rendezvous with Mars (elliptic orbit) is actually faster than the simplified circular coplanar rendezvous (0.977 years vs. 1.11 years). Comparison of this state history with the circular coplanar counterpart gives insights as to why this occurs (Figure 27). In the circular coplanar model, any mission start date is as good as any other therefore the boundary condition $q_0 = 0$ was a valid restriction where the Mars lead angle was determined in the numerical solution to the OCP. Because in the new events model we allowed the optimal q_0 to be determined in the OCP solution, a launch window was chosen such that Mars would be near the *slowest point* in its elliptic path at the time of rendezvous. Mars matches the sail velocity at the final event as it slows down on its approach to *aphelion.* The optimal rendezvous trajectory takes advantage of Mars' slow velocity as the sail approaches the planet at a radial distance that is 95% of the aphelion distance. Figure 28 shows how the sail meets the Mars position on its orbit.

Another interesting departure of the elliptic orbit solution from the circular solution is that the spacecraft does not follow a trajectory that sweeps out to a maximum radius and then returns inward to meet Mars at the required velocity (Figure 27). Mars' orbital path however does cross inside the spacecraft path as Figure 28 readily reveals. The optimal 3-D trajectory takes the fastest path from the optimal location on Earth's orbit to intercept Mars without having to undergo a negative radial velocity during the whole maneuver, i.e. $u > 0$ for all time (Figure 29).

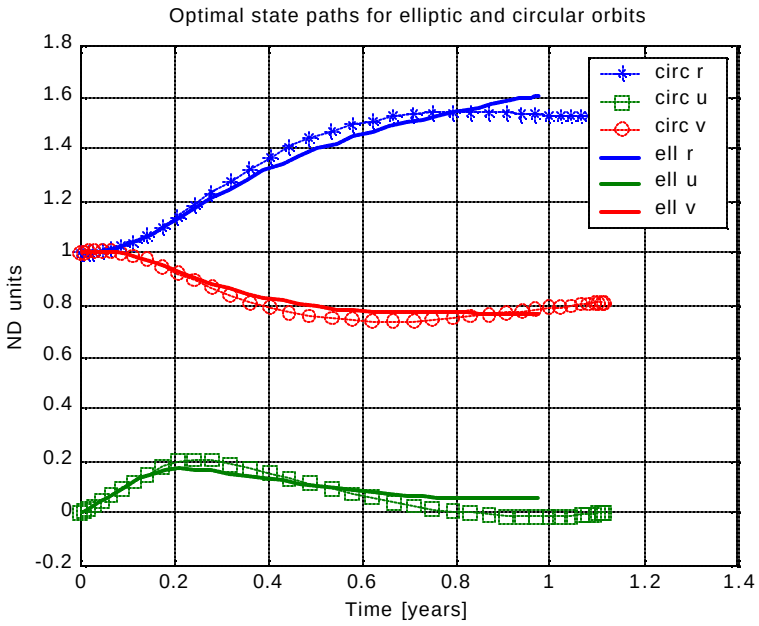

Figure 27 Comparison of State Profiles Using the Circular (markered lines) and Elliptic (thick lines) Orbits Models.

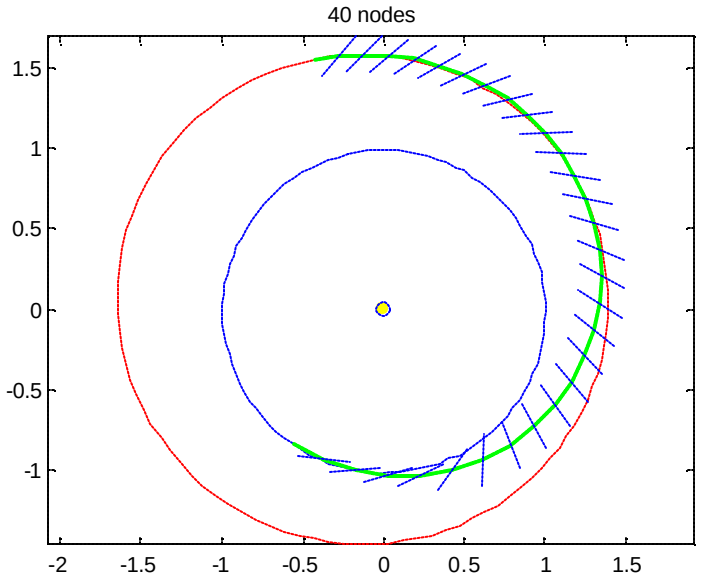

Figure 28 EM Rendezvous with Mars Elliptic Orbit

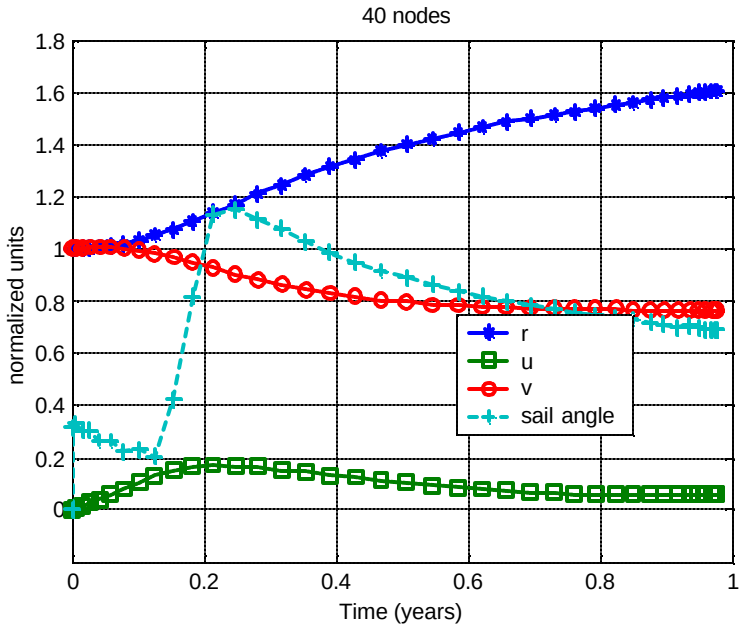

Figure 29 EM Rendezvous with Coplanar Elliptic Orbits.

c. Double Rendezvous

Once again, aphelion provides the optimal Mars rendezvous point (Figure 30 and Figure 1). The total transit time is only 2.3 years in contrast to the 2.4 years it took to reach the circular Mars orbit. In contrast to the double rendezvous orbits (see Figure 21), the radial velocity, u, does not change as much in the vicinity of the interior knot thus "flattening" the radial distance profile as the spacecraft sweeps past Mars.

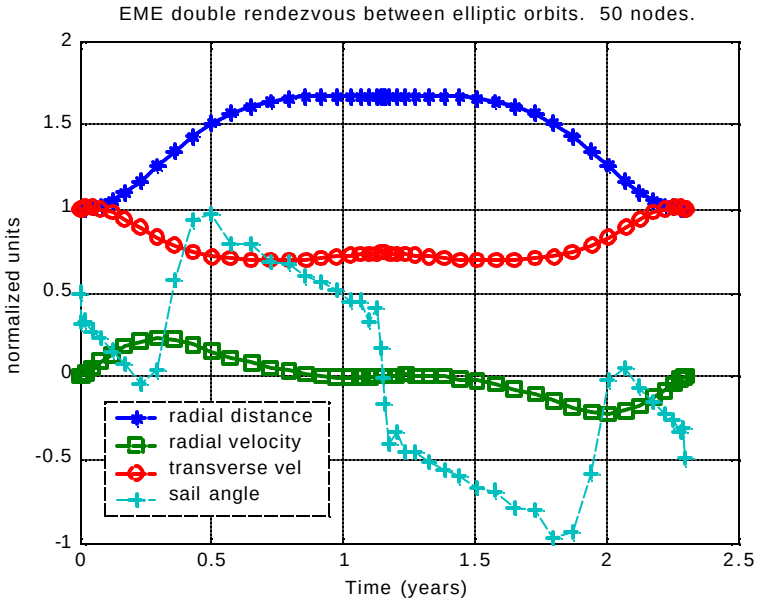

Figure 30 States and Control for EME Double Rendevous with Elliptic Mars Orbit.

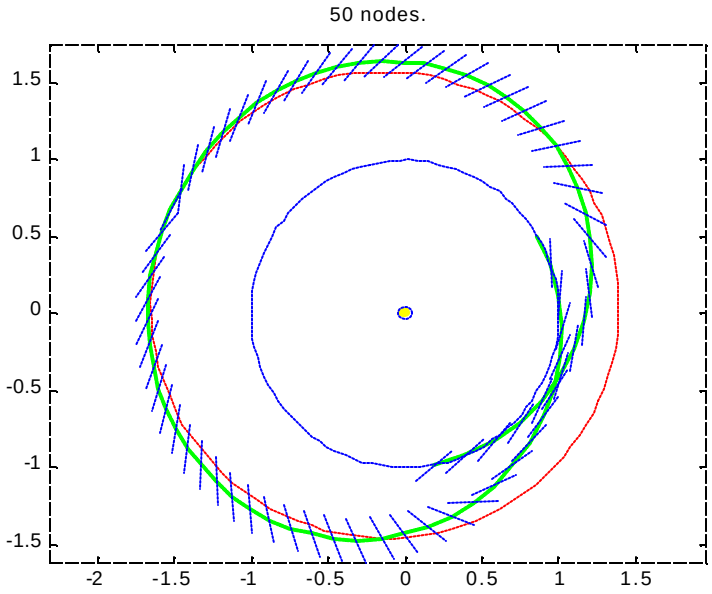

Figure 31 EME Double Rendezvous Trajectory and Sail Profile with Elliptic Mars Orbit.

3. Earth-Mars Synodic Cycler Solution

Having established a collection of solutions for basic solar sail trajectories, we are in a position to take the next step toward a cycler model. There are two key differences between the double rendezvous mission of the previous section and the solar sail cycler mission. With cyclers, end conditions are equal (i.e. cyclic) by definition to ensure repeatability. Additionally, gravity assists at target planets will be modeled. Planetary swingbys offer enhanced cycler performance since the turn angles are optimized to create a path that achieves the minimum cost. Since the time to complete a cycle is predetermined by the synodic period (Figure 32), a minimum time synodic cycler problem has no meaning for spacecraft shuttling between circular coplanar orbits. This is not so for the cycler between non-circular coplanar orbits. A new cost function is desired that is not burdened with minimizing fuel or time. Because an attractive feature of cycling trajectories is a slow swingby velocity at Earth and Mars, these velocities formed the cost function.

a. Events Model

The key to modeling a repeating cyclic trajectory is to constrain the initial state of the spacecraft to equal the final state of the spacecraft. Since the Earth and Mars orbits are approximated as circular and coplanar, the problem is simplified in that the initial relative angular position between Earth and Mars (lead angle z) may be used to constrain final angular position for a single cycle since planetary distances and velocities are independent of the their inertial angular positions (Figure 32). Lead angle is determined from the event states as follows.

$$q_{mf} = q_{m0} + n_m (t_f - t_0)$$

$$q_{ef} = q_0 + n_e (t_f - t_0)$$

where

$$q_{m0} = q_i - n_m (t_i - t_0)$$

Mars and Earth angular (q_m and q_e) are obtained using their respective mean motions (n_m and n_e).

The Earth-Mars lead angle at initial and final times are

$$z_0 = q_{m0} - q_{e0}$$
$$z_f = q_{mf} - q_{ef}$$

Initial and final relative angular positions are constrained by

$$\cos(z_f - z_0) = 1 \tag{13}$$

where $z_0 = z_f - 2p N_z$, $N_z = 0,1,2,...$

Numerically, equation (13) is preferable over the above equation since the cosine function will permit multiple revolutions without introducing integer variables.

The cyclic end condition for the spacecraft radial distance is expressed simply as

$$r_0 = r_f \tag{14}$$

Initial and final velocities are also constrained to be cyclic and will be addressed later.

Given the condition in equation (13), the final time, t_f, depends only on the relative mean motions of the planets given by the synodic period,

$$t_s = \frac{2p}{|n_e - n_m|}$$

For Earth and Mars, the synodic period is 2.135 years[13]. Although not explicitly constrained to the synodic period in the numerical analysis, the resulting final time must equal this synodic period in order for the cycler to be periodic. Having established cyclic end conditions, we now turn to setting up gravity assist conditions for planetary encounter events.

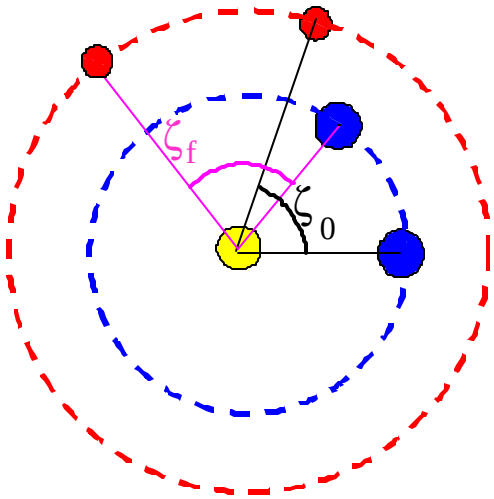

Figure 32 Two-Dimensional Earth-Mars Cycler Geometry with Circular Coplanar Planetary Orbits

Force imparted by the solar sail shapes the path *between* event manifolds, but gravity assist maneuvers *at* the event manifolds drive the form of the whole cycler trajectory. As with the conventional Aldrin cycler, gravity assists are implemented in the solar sail cycler to shape the trajectory at these planetary encounter events to improve the revisit times. To model the swingby events, state discontinuities are employed at the planet to change the velocity direction and magnitude of the spacecraft in the heliocentric frame (sometimes called the "zero sphere patched conic"[14] or matched asymptotes model). It is assumed that the interaction time with the subject planet is negligibly small in comparison to the total cycle time. The velocity of the spacecraft with respect to the planet changes direction such that $\mathbf{V}_{sp}^{+} = \mathbf{V}_{sp}^{-} + \Delta\mathbf{V}$, where $\mathbf{V}_{sp} = [u, v]^T - \mathbf{V}_p$ and $\mathbf{V}_p$ is the velocity of the subject planet relative to the sun. The position states, r and q are constrained to be continuous at both Earth (equation (14)) and Mars encounter events given below.

$$r^{-} = r^{+} \qquad q^{-} = q^{+} \tag{15}$$

The $\Delta\mathbf{V}$s due to the swingbys are optimally chosen in the OCP solution, but they must be properly constrained. Constraints on the velocity changes are most easily imposed using the *planet* frames where inbound and outbound velocity magnitudes are equal just before and after a planet encounter forming the event condition,

$$\left|\mathbf{V}_{sp}^{-}\right| = \left|\mathbf{V}_{sp}^{+}\right| = V_{\infty} \quad \text{(planet frame)} \tag{16}$$

where V_{∞} is the hyperbolic excess speed. The velocity direction change is expressed in the planetary frame using the turn angle, d, which exists in the region shown in . Velocities before and after a swingby event in the planet frame are coupled by the cosine of the turn angle,

$$\mathbf{V}_{sp}^{-} \bullet \mathbf{V}_{sp}^{+} = V_{\infty}^{2} \cos d$$

$$\cos d = \frac{\mathbf{V}_{sp}^{-} \bullet \mathbf{V}_{sp}^{+}}{V_{\infty}^{2}} \quad \text{(planet frame)} \tag{17}$$

The spacecraft experiences a direction change during the interaction that is restricted by the hyperbolic excess speed and the permissible periapsis pass distance from the center of the subject planet, r_p. This restriction is expressed by the following relationship in the planet frame where V_{∞} is scaled by the circular orbit speed at the surface of the planet and r_p is scaled by the radius of the planet (ref 15 p. 24).

$$\sin\left(d/2\right) = \frac{1}{V_{\infty}^{2} r_p + 1}, \; r_{p\min} \le r_p \le r_{p\max} \tag{18}$$

Although equations (17) and (18) themselves are not event conditions, they limit the achievable change in velocity direction, $\Delta\mathbf{V}$, due to the swingbys at Mars and Earth. Design limitations include a selected $r_{p\min}$ that is well above the atmosphere where drag effects are negligible and an $r_{p\max}$ such that the encounter occurs close enough to the planet to execute a desired task. For a given V_{∞}, the turn angle is maximum when $r_p = r_{p\min}$ and minimum when $r_p = r_{p\max}$. Substituting these values into equation (18) and solving for d as a function of V_{∞} provides an expression for $d_{\max}$ and $d_{\min}$ respectively.

To include the case of maximum deflection in the opposite direction it is necessary to place the lower bound of acceptable turn angles at $-d_{\min}$ and $-d_{\max}$ (Figure 33). With equation (17) characterizing the instantaneous change in path direction and equation (18) providing the limits, the boundary conditions are expressed at Earth and Mars event manifolds as

$$\cos(d_{\max}) \le \cos d \le \cos(d_{\min}) \tag{19}$$

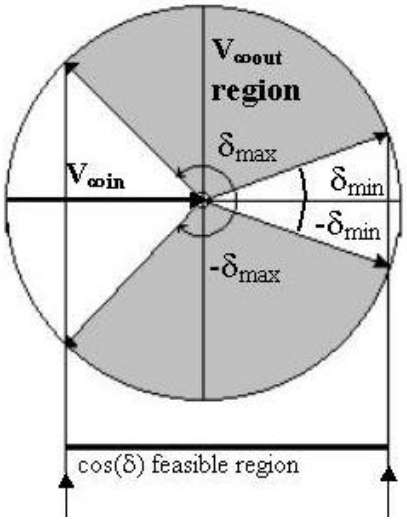

Figure 33 Gravity Assist Geometry.

Velocity constraints at Mars and Earth have identical form, however the initial velocity magnitude at Earth has an additional limitation based on available departure rocket capability. Because the sail's journey starts at Earth, the initial conditions are bounded by maximum C_3 available. Presumably an impulse rocket is used to start the solar sail craft on its cycler trajectory, so the initial velocity relative to Earth, $\mathbf{V}_0|_{earth}$, is restricted. The *magnitude* of $\mathbf{V}_0|_{earth}$ is limited by a maximum allowable velocity change at $t = 0$ provided by a kick motor, which provides another boundary condition

$$0 \le \left|\mathbf{V}_0|_{earth}\right| \le \Delta\mathbf{V}_{\max} \quad \text{(Earth frame)} \tag{20}$$

The direction of $\mathbf{V}_0|_{earth}$ is driven by the optimal control problem and is limited only by the allowable turn angle at each Earth swingby.

Finally, phasing the spacecraft with Earth at the end of a cycle was considered. Earth encounter events were constrained to ensure that the sail trajectory intersected Earth's orbital path at precisely the time that the planet is at that same location. The circular orbit assumption is particularly useful for ensuring proper phasing of events since the angular position of a planet is a linear function of time. The final angular position is given by

$$q_f = q_{e0} + n_e t_f - 2p N_s, \tag{21}$$

where N_s is an integer number of Earth orbits, n_e is Earth's mean motion, and q_{e0} is the angular position of Earth at $t = 0$ (when we the Earth position is coincident with the sail, $q_{e0} = q_0$). Having established the sail, dynamic and events models, we set up the optimal control problem.

b. Solar Sail Cycler Problem Formulation

The solar sail cycler optimal control problem is constructed using weighted spacecraft V_∞s at the Mars and Earth encounters as the cost function. The optimal path is subject to two-dimensional equations of motion, cyclic end conditions, and planetary gravity assist constraints. The cost function uses a parameter γ to weight the V_∞s while the initial and final state and control conditions are constrained to be equal to ensure repeatability of the trajectory. We can write the optimal control problem as the following.

Minimize the cost

$$J(\mathbf{x}_i, g) = gV_\infty|_{Mars} + (1-g)V_\infty|_{Earth} \tag{22}$$

Subject to dynamic constraints from the equations of motion (equation (2)) and event constraints that model repeatability (equations (13) and (14)), continuity

(equation (15)), swingby effects (equations (16) and (19)), launch limitations (equation (20)) and phasing (equation (21)).

In addition to the bounds on controls (equation (1)), there are bounds on states. States are bounded only to avoid singularities in the dynamics.

For this paper, a single Earth-Mars cycle was modeled where the following initial, intermediate and final conditions in astronomical units define the target manifolds

$$r_0 = r_f = 1$$
$$r_i = 1.524$$

The parameters that bound the path deflections at the event manifolds are the maximum C_3 available at the initial Earth orbit departure, $\Delta V_{\max}$ in canonical units, and the minimum and maximum periapsis pass distances at Earth and Mars.

$$\Delta V_{\max} = .2 \quad (\sim 6 \text{ km/s})$$
$$r_{m\min} = 1.06 \text{ Mars radii}$$
$$r_{m\max} = \textit{unbounded}$$
$$r_{e\min} = 1.16 \text{ Earth radii}$$
$$r_{e\max} = 10 \text{ Earth radii}$$

c. Synodic Cycler Results and Analysis

Shown in Figure 34 is the state and control angle output from DIDO for a single cycle of the synodic cycler with γ=1. As expected, the time required to complete a cycle under this set of constraints was 2.135 years, the Earth-Mars synodic period. A quick glance at the state history reveals that the spacecraft sails from 1 AU out to Mars' orbit at 1.524 AU and then returns to Earth. A discontinuity in both velocity states occurs at 1.524 AU and 1 AU representing Mars and Earth swingbys respectively. The optimal path makes use of large gravity assist maneuvers (in the planet frame) during planet encounters owing to slow V_∞s. The spacecraft initially accelerates in the radial direction while decelerating in the transverse direction. At the appropriate time, the sail rotates to an attitude that favors more and more positive transverse acceleration to intercept Mars with the lowest V_∞ to minimize the cost function while setting up for the swingby event

initiating the return trip. Following Mars swingby, the spacecraft sweeps out to nearly 2 AU to ensure proper phasing for Earth intercept. Sail attitude gradually reaches a maximum negative transverse acceleration profile (α = -35°; see equation (4)), then "shuts off" and follows a ballistic path as it presents an edge aspect to the sun. A plot of the sail trajectory with Earth 1, Mars 2 and Earth 3 encounters is shown in Figure 35. A similar gravity assist is accomplished at the Earth encounter and, since cyclic end conditions were imposed, the same control profile will reproduce the trajectory repeatedly. Owing to these constrained end conditions, the initial Earth departure hyperbolic excess velocity only required 4.3 km/s – not the maximum allowable limit of 6 km/s.

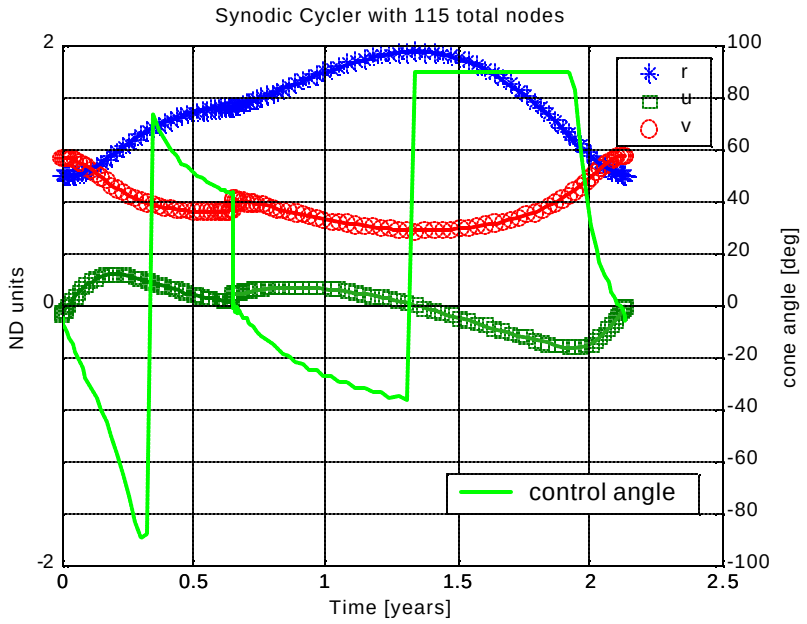

Figure 34 DIDO States (markers) and Control with Propagated Path (line through markers) for a Single Synodic Cycle.

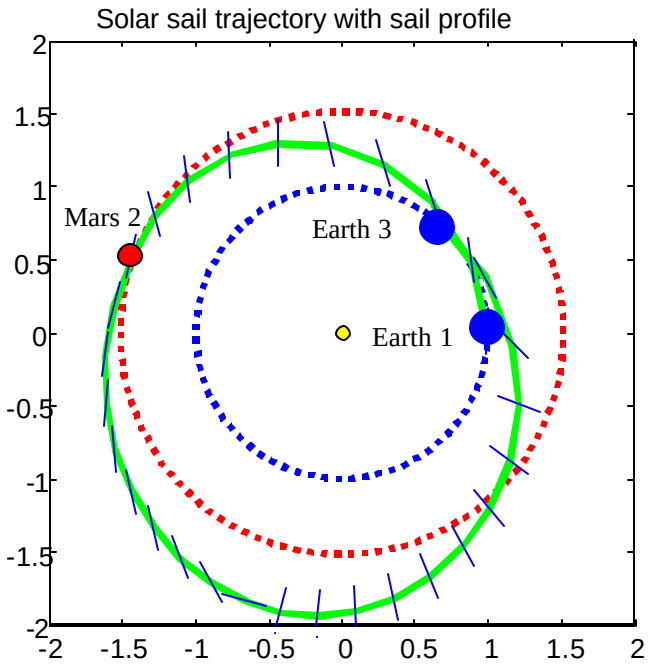

Figure 35 Single Cycle Path of Solar Sail Cycler with Minimum $V_{\infty}|_{mars}$.

A noticeable difference between the solar sail cycler and the traditional impulse rocket Aldrin cycler is the large swingby angular deflections with respect to Mars and Earth. The Aldrin cycler, because it is minimizing fuel mass, resides in a natural Keplarian orbit most of the time. As such, it tends to have a large V_{∞} in excess of 6 km/s at Mars and excess of 5 km/s at Earth, thus restricting turn angles. The solar sail, on the other hand, can change orbital energy with no impact to the cost function and achieve low hyperbolic excess speeds that permit large turn angles. The results of this analysis show that a 75° Mars turn angle and a 29° Earth turn angle provide the optimum path. Interestingly, the sail never goes down to the minimum allowable pass distance with Mars to get a bigger swingby deflection. Furthermore, the sail swings by Earth at the maximum allowable perigee distance, not the minimum allowable distance. Table 5 summarizes the cycler parameter data for the cost function with γ=1.

	V_{∞} *(km/s)*	r_p *(planet radii)*	*d*	*Time to planet*
Mars	2.53	1.27	75°	7.7 months
Earth	4.3	10	29°	18 months

Table 5 Earth-Mars Solar Sail Cycler Flyby Data (b=.17, g=1)

It should be noted that the parameters shown in Table 5 are highly sensitive to the states right before the Mars and Earth encounters. These numbers represent the optimal parameters given the approximated state and control history. By slightly modifying the approximations using more node points, states preceding a knot could change enough to generate a slightly different optimal parameter set.

To verify the solutions, states were propagated using initial conditions, gravity assist conditions and the DIDO-generated control history using a Matlab® ODE solver as described in the Validating Solutions section. Propagated states are shown passing through the DIDO output markers in Figure 34 producing the path in Figure 36.

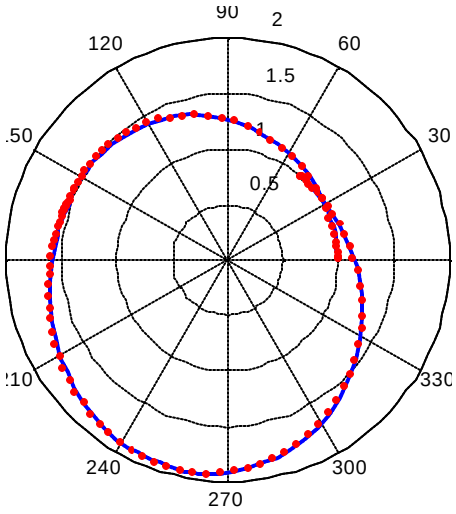

Figure 36 Propagated Path (line) with DIDO State Output (dots) for Single Cycle. Solution uses 115 total nodes (45 before and 70 after the interior knot).

To capture the effectiveness of the cyclic end condition constraint, the initial and final propagated states were compared for a single cycle. A comparison summary appears in Table 6 where the error between final and initial conditions of the propagated path are shown with states given as radial distance, r, Mars lead angle, z, velocity magnitude, V, and zenith angle, y (complement of flight path angle).

Final state	*Value*	*Initial state*	*Value*	*% Error*
r_f	1.0053	r_0	1.000	0.53
z_f	.6525	z_0	.6512	0.20
V_f	1.1249	V_0	1.1256	0.06
y_f	92.38°	y_0	93.86°	1.60

Table 6 Cyclic End Condition Errorsfor Synodic Cycler

Up to this point, we have only minimized the hyperbolic excess speed at Mars with the weighting factor g in equation (22) equal to unity. To suit the needs of any particular cycler mission, however, it may be desirable to minimize a combination of at both Earth and at Mars. Varying the weighting factor produces the range of V_∞s shown in the graph in Figure 37. A g of approximately 0.3 will minimize the cost function the most with least *total* V_∞.

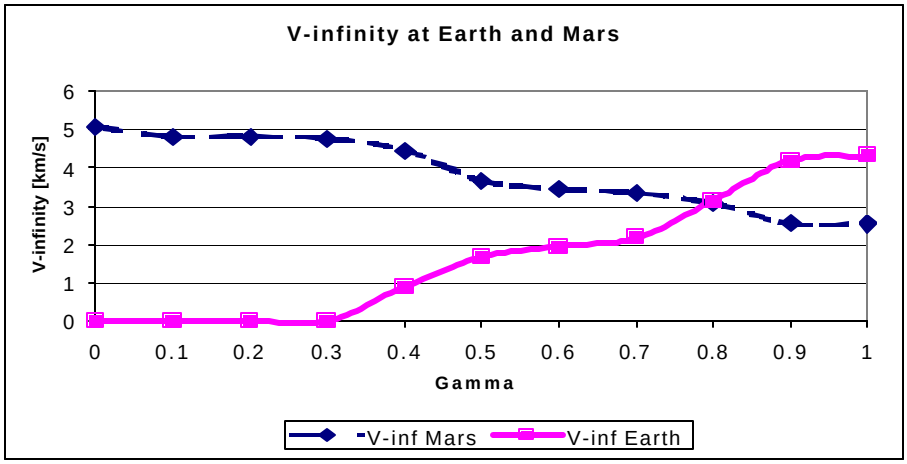

Figure 37 Varying g in the Complex Combination Cost Function

d. Remarks

Using a reasonably high performance solar sail to achieve optimal heliocentric cycler trajectories is a viable alternative to traditional impulse rocket cyclers. There are several interesting ways to pose an "optimal" solar sail cycler problem. One such problem would be a synodic cycler that achieves Mars and Earth encounters with the constraint that V_∞=0. This double rendezvous *synodic* cycler could also pose an intriguing design optimization problem in which one desires the minimum sail lightness number required to achieve a double rendezvous in a synodic period. These cases are investigated in the next section.

4. Fun with Cycler Trajectories

a. Double Rendezvous Synodic Cycler

Using the standard performance sail ($b = .17$), a cycler has been presented that minimizes V_∞ during planetary encounters. It may be interesting to seek a *synodic* cycler that is constrained to rendezvous with each planet such that V_∞ is zero at both planets. Recall that performing an EME double rendezvous mission with the standard sail with unrestricted end conditions took at least 2.41 years. This sail would never reach Earth again within a synodic period of 2.135 years. A higher performance sail however might be able to. Intuition tells us that there ought to be a sail with just barely enough area to achieve an EME double rendezvous within a synodic period. This sail design optimization problem may be stated as follows.

Find the minimum sail lightness number that would enable a double rendezvous cycler where $V_\infty = 0$ at both Earth and Mars.

Minimize: $J = b$

Subject to dynamic constraints

$$\dot{\mathbf{x}} - \mathbf{f}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$$

and event constraints given in Table 4 with the additional constraint that the end conditions are cyclic. Hyperbolic excess velocity is zero at the initial and final times, so the only additional constraint is that the EM lead angle, z, is the same at both times.

$$z_0 = z_f$$

The solar sail lightness number is established as a static parameter, a sail design characteristic that remains constant in time. The results in Figure 38 and Figure 39 show that the path has characteristics of both a synodic cycler and a double rendezvous. There is symmetry in the r and v states, and antisymmetry in the u state and control angle like the double rendezvous. Initial and final conditions are precisely the same as in a cycler, although no gravity assists are used. The time of flight turns out to be, of course, the EM synodic period. The minimum sail lightness number required to perform such a mission is β=.297. This corresponds to a sail that is nearly double the size of our standard solar sail with the same payload mass!

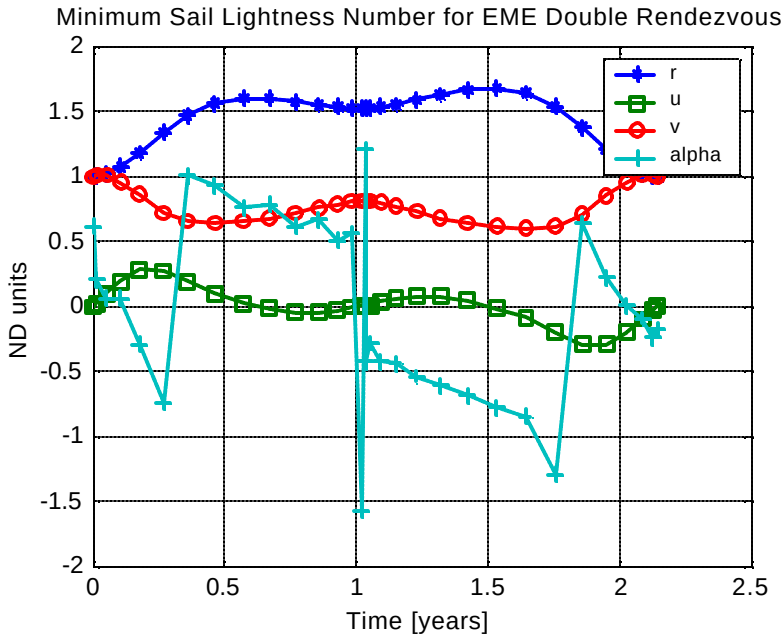

Figure 38 Minimum b Solar Sail States and Controls for an EME Double Rendezvous

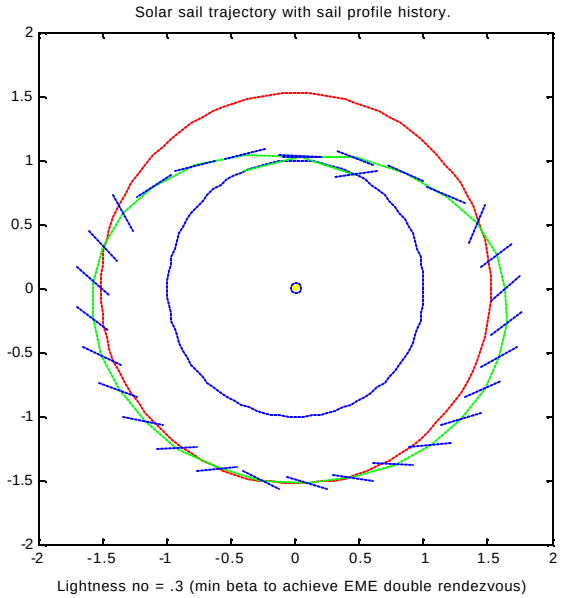

Figure 39 Trajectory for Minimum b EME Double Rendezvous

b. Taxi Propellant Cost

One mission that could benefit from a cycling orbit is replenishing supplies at a station on or around Mars. In concept, a taxi craft could leave its parking orbit about Mars and greet the passing cycler sail on its hyperbolic trajectory around Mars. A better cost function in this case would be the fuel required to meet the cycler on its swingby path from a parking orbit (equivalently, we could minimize the Δv). We assume a circular parking orbit with a radius equal to the closest cycler approach distance. Since the same exact pass conditions would be met every cycle, it is presumed that the taxi craft is positioned in this orbit. Following the discussion in ref 16 p. 102, the Δv required to go from a circular orbit to match the cycler's hyperbolic orbit is

$$\Delta v = \sqrt{\frac{m_m}{r_{pm}}} - v_{pm}$$

where r_{pm} and v_{pm} are the periapse position and velocity with respect to Mars. Since the energy of a hyperbolic orbit is

$$e_{hyp} = \frac{v_\infty^{\,2}}{2} = \frac{v_{peri}}{2} - \frac{m_m}{r_{pm}}$$

and the periapse velocity of the spacecraft with respect to Mars is

$$v_{pm} = \sqrt{v_\infty^{\,2} + \frac{2m_m}{r_{pm}}}$$

The change in velocity required to join the cycler craft on its hyperbolic trajectory is then

$$v_{pm} = \sqrt{v_\infty^{\,2} + \frac{2m_m}{r_{pm}}} - \sqrt{\frac{m_m}{r_{pm}}}$$

Note that if the circular parking orbit grazes the closest approach point of the cycler orbit (for fixed r_{pm}) the Δv is proportional to v_∞. Thus for this taxi model, a minimum $v_{\infty m}$ is equivalent to a minimum taxi Δv problem. For elliptic or 3D target orbit models where the closest cycler pass distances vary from visit to visit, the optimal taxi intercept path would differ. In this case, the cost would also have to include propellant expended to get from a nominal parking orbit to a non-grazing hyperbolic cycler path.

c. Profiles Using Different Sail Performances

As with any other blossoming technology, advances in sail material design are related to the amount of interest and thus funding. In order to make a cycler mission feasible it is useful to know what mission designs are available for any given sail performance. In the next analysis, a range of sail lightness numbers were fed to the dynamics model in equation (2) and analyzed in the same fashion presented using the standard sail. Results for $g = 1$ in the cost function of equation (22) for a range of lightness numbers are shown in Figure 40. The time to Mars, t_m, and $V_\infty|_{mars}$ are used to compare trajectory characteristics. The higher performance sails tend to take a longer time to reach Mars in order to reduce the hyperbolic excess speed at Mars.

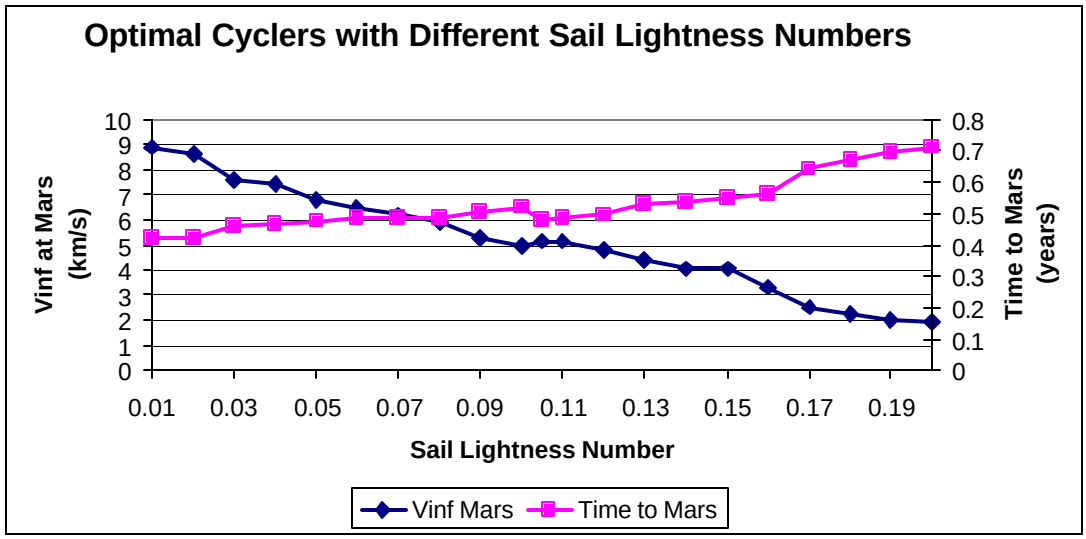

Figure 40 The Effect of Varying Sail Performances on a Cycler with g = 1 in the Cost Function

As the sail performance is reduced down to $b = .01$, the cycler flyby characteristics look more like a ballistic Aldrin cycler that passes closer to Earth using more bend angle. The sail attitude only changes the orbit enough to prevent the sail from dipping below the minimum pass distance restriction. A sail with 1/10 th the performance of the standard sail (i.e., β=.017) will reach Mars in a short time, get a small bend, then sweep outside Mars orbital path to return to Earth close enough to get a large bend. The planet encounter parameters are shown in Table 7.

	V_∞ *(km/s)*	r_p *(planet radii)*	*d*	*Time to planet*
Mars	8.4	2.9	7°	5.2 months
Earth	5.3	2.2	61°	20.4 months

Table 7 Solar Sail Cycler Characteristics for Cycler (b=.017, g=1).

E. VALIDATION OF SOLUTIONS

The primary method of validation was accomplished through numerical propagation. Using the same dynamic equations and initial conditions employed by DIDO along with the control history, a propagator will generate a trajectory that may be compared with the output state history. A match of resulting states indicates that the solar sail with the given control history follows a feasible path conforming to physical laws. The control profile may be obtained from the direct DIDO output, or derived from the costate history (for OCPs without an interior knot) and the sail control law. The latter generally provides "smoother" control for interior controls (when control angle is not at a limit, i.e. $-\frac{p}{2} < a < \frac{p}{2}$). A comparison of DIDO and costate-derived controls were shown in Figure 8for the flyby and Figure 15 for the rendezvous problem.

Numerical propagation was accomplished using the Matlab® ODE45 and ODE23 solvers. Controls had to be interpolated between DIDO node points in order to produce an accurate sail attitude at time steps generated by the ODE solver. Generally, a cubic interpolation served well while a spline method proved inaccurate in regions with concentrated node points, i.e. near knots.

Validations of the benchmark problems are shown in Figure 9 and Figure 17. DIDO paths and the propagated paths closely match as shown in Table 8 listing the mean squared error for the different runs (DIDO dynamic constraints were met with less than 10^{-5} for both runs).

State	***Flyby***	***Rendezvous***
r	1.2393e-009	1.0944e-006
q	6.3148e-009	3.3749e-004
u	1.0952e-009	3.4343e-006
v	1.0258e-009	3.4093e-006

Table 8 Mean Square Error of DIDO States Compared with Propagated States Using Matlab® ODE45

Propagation of swingby trajectories, such as those used in the cycler problems, became more of an art than a science. The propagation was enacted for one cycle only with an intermediate event condition defined at Mars' orbital radius. To simulate the gravity assist, a switch was used in the propagator to add a ΔV to the sail's velocity prior to the encounter equal to the ΔV determined by DIDO. Because of the sharp discontinuity in the path velocity due to the gravity assist at the knot, the subsequent trajectory was driven by where the assist occurred. Small differences in where the outbound path encountered Mars caused the DIDO states and propagated results to differ on the inbound leg Figure 41. The discrepancy lay in how the sail was being controlled just prior to the interior knot. Output controls often appear "shaky" near a knot. Wit hin the propagator, these somewhat erratic controls must be approximated at the time steps. If the time steps of the propagator were large in comparison to the DIDO controls at the LGL points, approximation errors resulted. These small errors are enough to slightly change the Mars swingby event location causing the remainder of the trajectory to deviate from the DIDO solution. The discrepancy is resolved by forcing the propagator step sizes to be approximately the same size as the distance between DIDO-generated LGL points near the knot (Figure 42). In this way, better approximations are made near the defined points producing closer matches between DIDO and propagated paths. This can be accomplished by either imposing a maximum step size on the propagator or by adjusting the number of nodes used by DIDO to manage node spacing. Better control approximations in the neighborhood of the planetary encounters yielded a closer match between the optimal solution states and the propagated states. This, however, only confirms that the output really does match a propagated solution for the given erratic control behavior just before the hard knot. We desire a solution that does not require the sail to perform radical attitude changes near the Mars encounter.

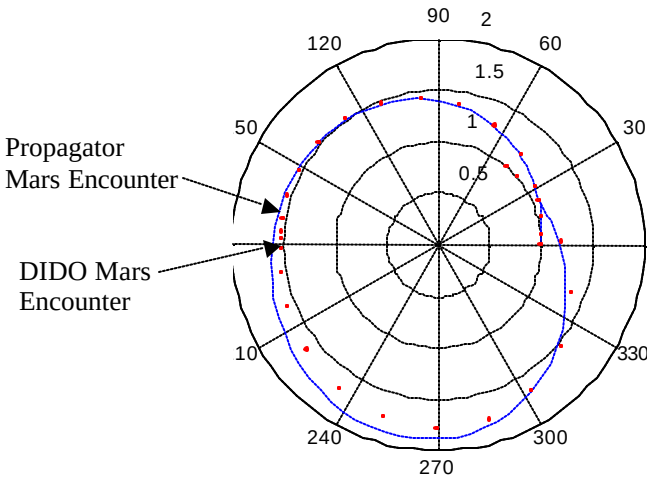

Figure 41 Propagated Path with "Missed" Mars Swingby Phasing. Propagated states use controls interpolated at time steps different than DIDOs LGL distributed time steps causing differences at the interior hard knot.

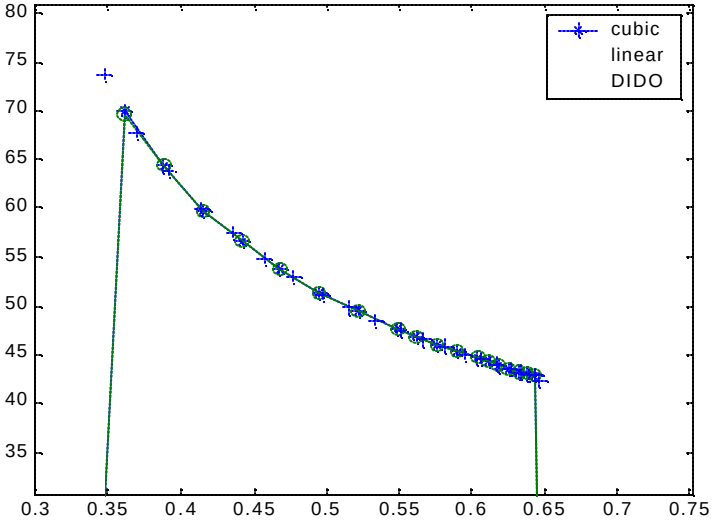

Figure 42 DIDO Control Output at LGL Node Points and Interpolated Controls used in the ODE45 Propagation Near a Knot. Step sizes match fairly well.

The controls may be made “smoother” by adding “inertia” to the controls; i.e. by limiting the rates. The control then becomes the time rate of change of the cone angle.

$$\mathbf{x} = [r \quad \theta \quad u \quad v \quad \alpha]^T$$

where the control is $u_a = \dot{\alpha}(t)$.

Using this control and state variable convention provided inherently stable cone angle control, a more desirable design for a solar sail attitude control system. Additionally, the pitch rate, $\dot{\alpha}$, was be restricted to provide the “inertia” to the sail without having to switch from a 3 DOF problem to a 4 DOF problem. Numerical solution time was expected to shorten as well since the hodograph of “control” $\dot{\alpha}$ is convex (see Appendix E).

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DEVELOPING THE OPTIMAL THREE-DIMENSIONAL CYCLER

We now extend the boundary conditions and dynamics of the 2-D solar sail trajectory models of the previous chapter to include a third dimension and use the same methodical approach to solve the optimal cycler. The definition of a "cycle" for the 3 dimensional problem however becomes more complex than the 2-D case. Initial and final cycle end conditions do not repeat exactly in a synodic period as they did in the 2-D circular coplanar model. Planetary orbit inclination and eccentricity make conditions such that Earth and Mars do not repeat their relative positions with each other but about every 15 years.

The motivation for obtaining a 3-D optimal control profile for a solar sail cycler is not only to increase the fidelity of the model but also to compare results to the 2-D case to learn more about the nature of optimal cycler orbits. The inclination of Mars' orbit with respect to the ecliptic plane is roughly 1.85°, not a great difference from the coplanar case. Elliptic orbits for Earth and Mars are modeled, so optimal trajectories are similar to the elliptic coplanar solutions. We perform the same series of benchmark 3D solutions to rate them against their 2D circular and elliptic counterparts.

A. THREE-DIMENSIONAL MODELS

1. Sail Model

Our sail in the new model now has an extra degree of freedom due to the addition of the third dimension. With most conventional engines, the controls generally have three degrees of freedom, one for each of three dimensions. However, with the sail, thrust magnitude is dependent on the cone angle providing a constraint, therefore only two degrees of freedom are required. The cone angle serves well as one of the control variables since the thrust magnitude is related to it. Another angle is required to determine where on the cone the sail normal vector lies. It is convenient to define a clock angle as the angle between the projection of the sail normal onto a plane normal to the

sun-line and a reference direction in the same plane; so we adopt a model used in ref 9 p.115. This reference direction is taken to be the vector normal to the instantaneous orbital plane of the sail. Figure 43 makes the representation more clear with unit vector $\hat{\mathbf{p}}$ as the reference direction. Note that a positive clock angle rotation is in the negative $\mathbf{r}$ direction using the right hand rule.

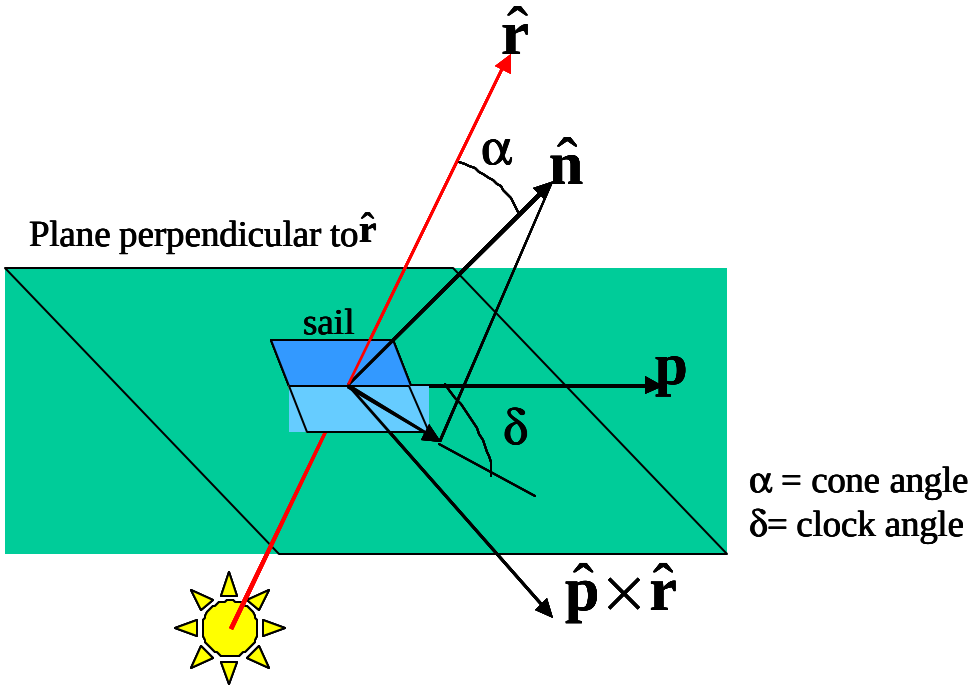

Figure 43 Solar Sail Control Model for 3 Dimensional Dynamics

2. Dynamics Model

Since the thrust magnitude is dependent on the radial distance from the sun, spherical equations of motion provide a simple way of including the acceleration due to the sail (see Coordinate Systems Section)

For the three degree of freedom state variable $\mathbf{x} = [r\ q\ f\ v_r, w_q, w_f]^T$ the state derivative is given by

$$(23)\quad \mathbf{f} = \begin{bmatrix} v_r \\ w_q \\ w_f \\ r w_q{}^2 \cos^2 f + r w_f{}^2 - \frac{m}{r^2} + \frac{bm}{r^2}\cos^3 a \\ \frac{1}{r\cos f}\left[2w_q \left(r w_f \sin f - v_r \cos f \right) + \frac{bm}{r^2}\cos^2 a \sin a \sin d \right] \\ \frac{1}{r}\left(-2 v_r w_f + \frac{bm}{r^2}\cos^2 a \sin a \cos d \right) - w_q{}^2 \sin f \cos f \end{bmatrix}$$

where the last three terms are from the equations of motion (Appendix C).

In choosing these dynamic equations, we must bound the states to avoid singularities at $r = 0$ and $\cos f = 0$. It turns out that avoiding these state values keeps the Jacobian of $\mathbf{f}$ free of singularities as well.

3. Events Model

The 3D target orbit manifolds are the locations and corresponding velocities of Earth and Mars along their respective inclined elliptic orbits. These events define the boundary conditions of the optimal control problem. In defining the boundary conditions, it is necessary to know the relative orbital shapes and orientations of the departure and destination planets. The orbital elements of Earth and Mars orbits are summarized in the following table[13].

	a [AU]	***e***	***i***	*W*	*w*
Earth	1.0	.0167	0	undefined	102.9°
Mars	1.524	.0935	1.85°	49.57°	286.5°

Table 9 Earth and Mars Orbital Parameters

In all the problems presented, the planet's true anomaly at the events remains a free variable to be determined in the numerical solution.

In order to rendezvous with a planet, it is essential to describe the position and velocity of both spacecraft and planet in a common 3-D (N_D=3) coordinate system. Although the spacecraft states are represented in spherical coordinates, the events at t_0 and t_f are given in perifocal coordinates in the respective planetary orbital planes. To match the sail states with the planetary states at the end conditions, Earth and Mars orbital planes were transformed into a common heliocentric-ecliptic coordinate system along with their respective velocities along the orbital paths. For simplicity, the frame is referred to as the E-frame. These orbital states define the target manifolds to which the spacecraft event conditions are to be constrained (see Appendix D).

First, the initial Earth angular position and final Mars angular position are constrained to be equal to the sail's position in the common E-frame. Resolving these manifolds in the E-frame we obtain the planetary positions in a common coordinate system.

Constraining the initial and final sail angular positions to be coincident with the initial Earth and final Mars angular positions respectively, we obtain for small inclinations (ref 4 p. 135) in the E-frame

$$q_{e0} = q_0 - (w_e + \Omega_e) \qquad\qquad q_{mf} \approx q_f - (w_m + \Omega_m)$$

where θ_0 is the initial spacecraft position and θ_f is the final sail position.

In a perifocal system, the planetary polar coordinates r_p and θ_p are related by

$$r_{e0} = \frac{p_e}{1 + e_e \cos q_{e0}} \qquad\qquad r_{mf} = \frac{p_m}{1 + e_m \cos q_{mf}}$$

where p in the numerator is the "parameter" or semi-latus rectum of the respective orbit.

Now the initial and final sail target positions in Cartesian coordinates are given by

$$\mathbf{R}_{eE} = \mathbf{A}_e \begin{pmatrix} r_{e0}\cos(q_0) \\ r_{e0}\sin(q_0) \\ 0 \end{pmatrix} \text{ for Earth and } \mathbf{R}_{mE} = \mathbf{A}_m \begin{pmatrix} r_{mf}\cos(q_f) \\ r_{mf}\sin(q_f) \\ 0 \end{pmatrix} \text{ for Mars}$$

where $\mathbf{A}_e$ and $\mathbf{A}_m$ are the respective Earth and Mars 3-1-3 rotations transforming into the E-frame (Appendix D). In like manner the velocities of the planets may be defined in their own perifocal frame by

$$V_p = \sqrt{m\left(\frac{2}{|\mathbf{r}_p|} - \frac{1}{a_p}\right)}$$

where V_p is the speed of the planet, $|\mathbf{r}_p|$ is the distance of the planet from the sun, and a_p is the semi-major axis of the planet. Components of the planetary velocity vector are defined in terms of the flight path angle b by

$$\cos b_p = \frac{h_p}{|\mathbf{r}_p| V_p} \text{ and } \sin b_p = \frac{m}{h_p V_p} e_p \sin\left(q_p\right)$$

where h_p represents the magnitude of the planet's orbital angular momentum vector.

A planet's velocity vector resolved in the local vertical local horizontal frame (LVLH) has components

$$u_p = V_p \sin b_p$$

$$v_p = V_p \cos b_p$$

Finally resolving these components in Cartesian coordinates and transforming them into the E-frame yields the target velocities $\mathbf{V}_{eE}$ and $\mathbf{V}_{mE}$ (transformation is given in Appendix D). For the rendezvous mission, $\mathbf{V}_{mE}$ is required to match the final sail velocity, but knowledge of Mars' velocity is not needed for the flyby mission.

The constraints may now be set such that the spacecraft states at the initial and final events are equal to the target planetary states. The initial and final sail positions are given as follows.

$$\mathbf{R}_{sE} = \begin{pmatrix} x_0 & x_f \\ y_0 & y_f \\ z_0 & z_f \end{pmatrix}$$

The left column represents the sail position at the beginning, $\mathbf{r}_0$, and the right column represents the sail position at the final time, $\mathbf{r}_f$. All coordinates are resolved in the E-frame. After applying the transformation given in equation (45), the spacecraft velocities at events in LVLH frame are given by

$$\mathbf{V_s} = \begin{pmatrix} v_{r0} & v_{rf} \\ v_{q0} & v_{qf} \\ v_{f0} & v_{ff} \end{pmatrix}$$

The velocities need to be transformed from LVLH into Cartesian coordinates in the E-frame using a transformation matrix $\mathbf{B}$ defined as

$$\mathbf{B} = \underset{2\to 3}{[B]} = [R_3(-q)][R_2(f)]$$

where the R matrices correspond to standard rotation matrices (Appendix D). The spacecraft velocity in the Cartesian E-frame is therefore

$$\mathbf{V}_{sE} = \mathbf{B}\mathbf{V}_s, \quad V_{sE} \in \mathbb{R}^{N_D \times N_E}$$

The left column of $\mathbf{V}_{sE}$ represents the initial sail velocity, $\mathbf{V}_0$, and the right column of $\mathbf{V}_{sE}$ represents the final sail velocity, $\mathbf{V}_f$. Having defined all the necessary states in the same coordinate sys tem, it is now a simple matter to set boundary conditions for the initial and final rendezvous events.

The sail position starts at Earth and ends at Mars

(24) $$\boxed{\mathbf{R}_0 - \mathbf{R}_{eE} = \mathbf{0}}$$

(25) $$\boxed{\mathbf{R}_f - \mathbf{R}_{mE} = \mathbf{0}}$$

and likewise the initial sail velocity is the orbital velocity of Earth and final velocity is equal to the orbital velocity of Mars.

(26) $$\boxed{\mathbf{V}_0 - \mathbf{V}_{eE} = \mathbf{0}}$$

(27) $$\boxed{\mathbf{V}_f - \mathbf{V}_{mE} = \mathbf{0}}$$

The final velocity event constraint in equation (27) applies to the rendezvous mission only where the spacecraft must match the velocity of Mars in its orbit. Table 10 summarizes the event conditions in equations (24) through (27) for a rendezvous. All coordinates are resolved in Cartesian coordinates in the E-frame. Within the DIDO framework we code the event conditions in a more compact form that ensures end condition manifolds are constrained.

$$\mathbf{R}_{sE} - \mathbf{R}_{eE} = \mathbf{0} \text{ and } \mathbf{V}_{sE} - \mathbf{V}_{eE} = \mathbf{0}$$

For only two end conditions without intermediate conditions the dimensions of all the matrices are $N_D \times N_E$, where N_D is the number of dimensions and N_E is the number of events.

Event	***Flyby constraints [lower,upper]***	***Rendezvous constraints [lower,upper]***
$\mathbf{R}_0$	$[\mathbf{R}_{eE}, \mathbf{R}_{eE}]$	$[\mathbf{R}_{eE}, \mathbf{R}_{eE}]$
$\mathbf{V}_0$	$[\mathbf{V}_{eE}, \mathbf{V}_{eE}]$	$[\mathbf{V}_{eE}, \mathbf{V}_{eE}]$
$\mathbf{R}_f$	$[\mathbf{R}_{mE}, \mathbf{R}_{mE}]$	$[\mathbf{R}_{mE}, \mathbf{R}_{mE}]$
$\mathbf{V}_f$	*free*	$[\mathbf{V}_{mE}, \mathbf{V}_{mE}]$

Table 10 Event Conditions for 3-D Flyby and Rendezvous Missions.

B. THE OPTIMAL CONTROL PROBLEM

In keeping with the step-by-step approach to reaching a cycler model, we first seek the minimum time solutions to the 3-D flyby and rendezvous missions. To this end the new 3D OCP is stated in a similar fashion as the 2D OCP. Because the state variable has now expanded to six variables (3 position, 3 velocity), we constrain the state

derivatives to the 3D equations of motion (equation (23)). States are bounded to avoid the additional singularity at the solar poles.

C. SOLAR SAIL CONTROL LAW

Since the sail attitude control system is responsible for driving both the cone angle, α, and clock angle, δ, we seek two steering laws. As with the 2-D model, we turn to the principles of optimal control theory to generate these steering laws. Tailoring the Bolza cost function in Lagrange form for a minimum time problem we get

$$J = E + \int_{t_0}^{t_f} F dt$$

$$E = 0, F = 1$$

Using states consistent with a spherical coordinate system, $\mathbf{x} = [r, q, f, v_r, w_q, w_f]^T$, we construct our state derivative vector $\mathbf{f}$ from the equations of motion (Appendix C).

$$(28)\ \mathbf{f}(\mathbf{x}, a, d) = \begin{bmatrix} v_r \\ w_q \\ w_f \\ r w_q^{\,2} \cos^2 f + r w_f^{\,2} - \frac{m}{r^2} + \frac{bm}{r^2} \cos^3 a \\ \frac{1}{r \cos f} \left[2 w_q \left(r w_f \sin f - v_r \cos f \right) + \frac{bm}{r^2} \cos^2 a \sin a \sin d \right] \\ \frac{1}{r} \left(-2 v_r w_f + \frac{bm}{r^2} \cos^2 a \sin a \cos d \right) - w_q^{\,2} \sin f \cos f \end{bmatrix}$$

Thus, the Hamiltonian is

$$H(\mathbf{x}, \mathbf{u}, \lambda) = F + \lambda^T \mathbf{f}(\mathbf{x}, \mathbf{u})$$

$$H = 1 + l_r \dot{r} + l_q \dot{q} + l_f \dot{q} + l_{vr}\left(r\dot{q}^2 \cos^2 f + r\dot{f}^2 - \frac{m}{r^2} + \frac{bm}{r^2}\cos^3 a \right)$$

$$+ \frac{l_{vq}}{r\cos f}\left[2\dot{q}\left(r\dot{f}\sin f - \dot{r}\cos f \right) + \frac{bm}{r^2}\cos^2 a \sin a \sin d \right]$$

$$+ \frac{l_{vf}}{r}\left(\left(-2\dot{r}\dot{f} + \frac{bm}{r^2}\cos^2 a \sin a \cos d \right) - \dot{q}^2 \sin f \cos f \right)$$

The Lagrangian is therefore

$$L(\mathbf{x}(t), \mathbf{u}(t), \mathbf{l}(t)) = H(\mathbf{x}(t), \mathbf{u}(t), \mathbf{l}(t)) + \boldsymbol{\mu}_g^{\,T}(t)\mathbf{g}(\mathbf{x}(t), \mathbf{u}(t))$$

or simply,

$$L = H + m_d d + m_a a \text{ where } a \in \left[-\frac{p}{2}, \frac{p}{2} \right] \text{ and } \mu_\delta = 0 \text{ if } \delta \text{ is unconstrained.}$$

However in the actual implementation of control bounds, δ is limited by $d \in \left[-\frac{p}{2}, \frac{p}{2} \right]$ to avoid non-distinct controls where two (α, δ) control pairs could produce the same resulting thrust vector. It will be apparent later why this restriction is important.

First deriving the control law for the clock angle we calculate

$$\frac{\partial L}{\partial d} = \frac{l_{vq}}{r\cos f}\frac{bm}{r^2}\cos^2 a \sin a \cos d - \frac{l_{vf}}{r}\frac{bm}{r^2}\cos^2 a \sin a \sin d + m_d = 0$$

Dividing by $\frac{bm}{r^3}\cos^2 a \sin a$, we obtain the tangent steering law for the clock angle δ interior controls for $a \neq 0, \pm\frac{p}{2}$.

(29) $$\boxed{\tan d = \frac{l_{vq}}{l_{vf}\cos f}},\ m_d = 0$$

The control angle δ has no effect on the resulting thrust vector when the sail is exposing its full area to the sun or when it is exposing no area $\left(a = 0, \pm\frac{p}{2} \right)$. Applying

the KKT conditions (Appendix A) to obtain the circumstances when the clock angle is at the "stops" we get

$$\delta = -\frac{\pi}{2}, \mu_\delta \le 0$$

$$\delta = +\frac{\pi}{2}, \mu_\delta \ge 0$$

where the sign of the covector determines which way the clock angle is oriented.

Next, we desire the steering law for the cone angle, α. Applying the same KKT conditions to cone angle, α, we obtain similar results. The "stops" of the cone angle control occur at $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ when the control dual variable meets the following conditions.

$$\alpha = -\frac{\pi}{2}, \mu_\alpha \le 0$$

$$\alpha = +\frac{\pi}{2}, \mu_\alpha \ge 0$$

The "interior" cone control is defined where the dual variable is zero.

$$-\frac{\pi}{2} < \alpha < \frac{\pi}{2}, \mu_\alpha = 0$$

Continuing with the Minimum Principle, we obtain

$$\frac{\partial L}{\partial \alpha} = -\lambda_{vr}\frac{\beta\mu}{r^2}\left(3\cos^2\alpha\sin\alpha\right) + \frac{\lambda_{vq}}{r\cos\phi}\frac{\beta\mu}{r^2}\sin\delta\left(-2\cos\alpha\sin^2\alpha + \cos^3\alpha\right)$$

$$+\frac{\lambda_{vf}}{r}\frac{\beta\mu}{r^2}\cos\delta\left(-2\cos\alpha\sin^2\alpha + \cos^3\alpha\right) + \mu_\alpha = 0$$

Dividing by $\frac{\beta\mu}{r^2}\cos^3\alpha$ produces

$$(30)\quad -3\lambda_{vr}\tan\alpha + \frac{\lambda_{vq}}{r\cos\phi}\sin\delta\left(-2\tan^2\alpha + 1\right) + \frac{\lambda_{vf}}{r}\cos\delta\left(-2\tan^2\alpha + 1\right) + \mu_\alpha = 0$$

where $a \neq \pm\frac{p}{2}$. This corresponds to the *interior* controls only when the cone angle multiplier is zero, $m_a = 0$. Using equation (29) for tanδ, we form an expression for $\cos\delta$ and $\sin\delta$ as a function of the states and costates.

$$\cos d = \pm\frac{l_{vf}\cos f}{\sqrt{\left(l_{vf}\cos f\right)^2 + l_{vq}^2}} \qquad \sin d = \pm\frac{l_{vq}}{\sqrt{\left(l_{vf}\cos f\right)^2 + l_{vq}^2}}$$

The sign ambiguity is resolved when we take into account the control bounds, i.e. $d \in \left[-\frac{p}{2}, \frac{p}{2}\right]$, therefore $\sin d > 0$. Recall that equation (29) is not valid at $a = 0$.

Rearranging equation (30) into standard quadratic polynomial form for tanα

$$a\tan^2 a + b\tan a + c = 0, \quad a \in \left(-\frac{p}{2}, \frac{p}{2}\right) \text{ and } a \neq 0$$

where the coefficients are given by the following:

$$a = -2\frac{l_{vq}}{r\cos f}\sin d - 2\frac{l_{vf}}{r}\cos d$$

$$b = -3\,l_{vr}$$

$$c = \frac{l_{vq}}{r\cos f}\sin d + \frac{l_{vf}}{r}\cos d$$

Solving the quadratic yields the tangent steering law for the cone angle interior controls.

(31) $$\boxed{\tan a = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}} \text{ for } a \in \left(-\frac{p}{2}, \frac{p}{2}\right) \text{ and } a \neq 0$$

Following numerical solutions to simple optimal control problems (i.e. no interior knots), DIDO will output the time histories of costate as well as dual variables associated with the control angles. Steering laws in equations (29) and (31) use these outputs to

generate control profiles that may be compared to the DIDO control history output for verification.

D. RESULTS

1. Benchmark Problem Solutions with *Elliptic Inclined* Orbits

The question in reviewing the same series of benchmark problems with elliptic inclined orbits is the following. Does adding a degree of freedom in the dynamics to meet the 3D event conditions enable us to further reduce the cost from the elliptic coplanar case or does it hurt us? Apparent from the first several solutions is that the 3 DOF model minimizes the transit time to a smaller value than the circular coplanar model, but produces a larger trajectory time than the elliptic coplanar case.

a. Earth-Mars Flyby

To reach Mars as quickly as possible and sail past without gravitational interaction, it was intuitive that Mars periapsis provided the minimum path distance in the coplanar model. Now when we consider that the target orbital planes are inclined with respect to each other it may be desirable to choose a path that avoids orbit "cranking" even when hitting the periapsis is not possible in plane. The solution shows that the sail changes planes a small amount in order to reach Mars at its perihelion as shown in Figure 44. Mars' orbit has a small inclination with respect to the ecliptic so it is preferable to traverse a short distance even though a small cranking maneuver is incurred. Figure 45 reveals that the time to intercept Mars with our standard sail (β=.17) is 138 days, which is about the same as transiting to an in-plane elliptic Mars orbit (137 days). The sail cone angle control history is very similar to the coplanar case cone history while the clock angle steadily increases (Figure 46). Controls derived from the tangent steering law using costates are compared to the DIDO control output showing a close match.

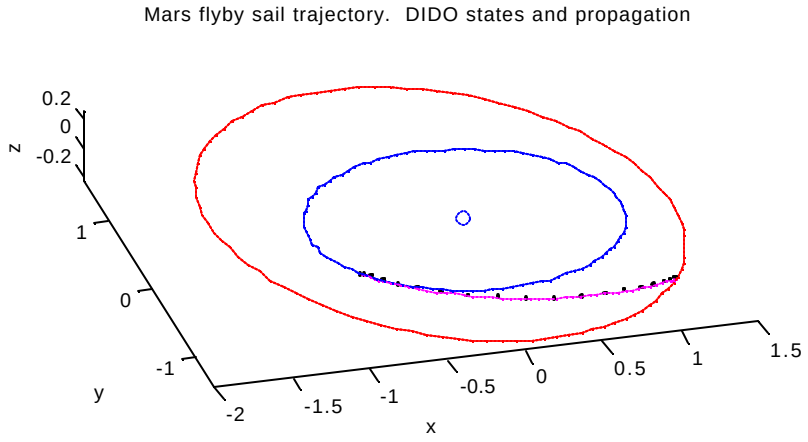

Figure 44 Flyby Mission to Mars with Elliptic, Inclined Planetary Orbital Planes. DIDO output (dots) and propagated path (line). Mars orbit inclination is exaggerated for display purposes.

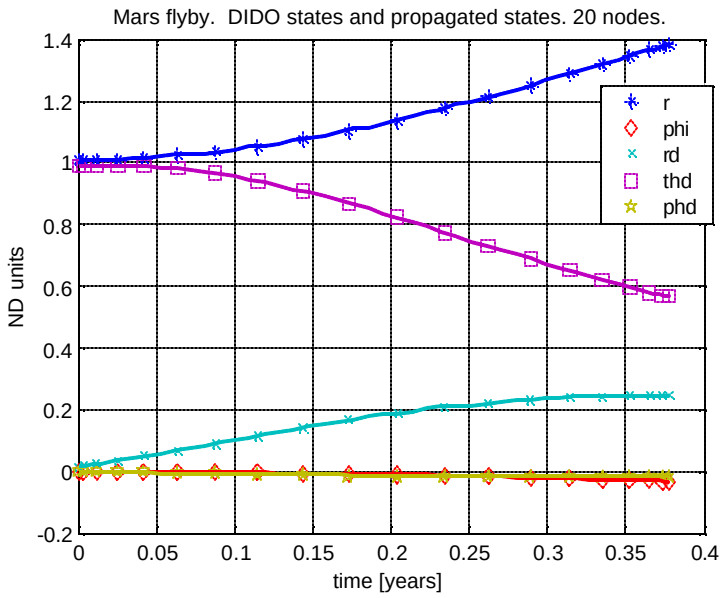

Figure 45 State History for Mars Flyby (3D Model)

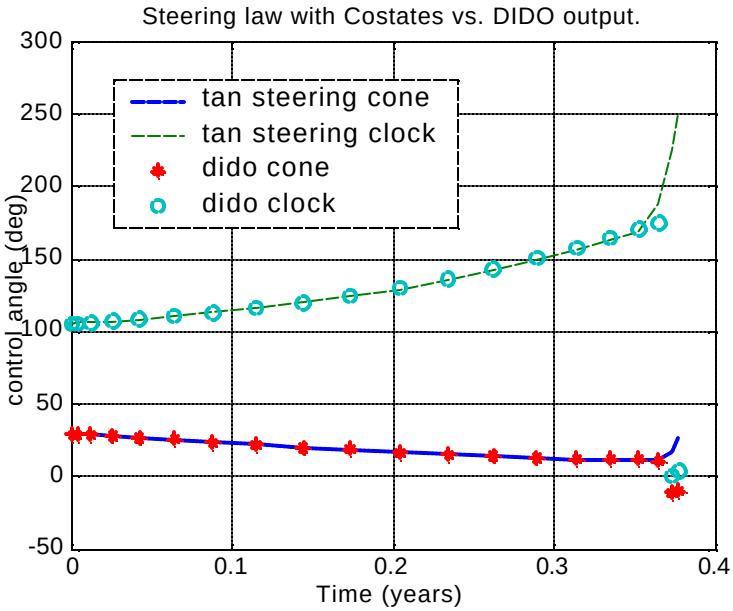

Figure 46 Cone and Clock Angle Controls for Mars Flyby. History is shown for DIDO (markers) and tangent steering law using states and costates (lines).

b. Rendezvous

Observing the f and w_f states in Figure 47 and Figure 48 gives insight into what is required to perform the plane change maneuver. The sail trajectory remains in the Earth ecliptic plane for the first quarter of the mission time, then "cranks" the orbital plane to match the final sail f and w_f with Mars exactly at the time of Mars encounter. The mission is 1.01 years, which is a month faster that the circular coplanar and takes only marginally longer than the elliptic coplanar orbit model (0.977 years).

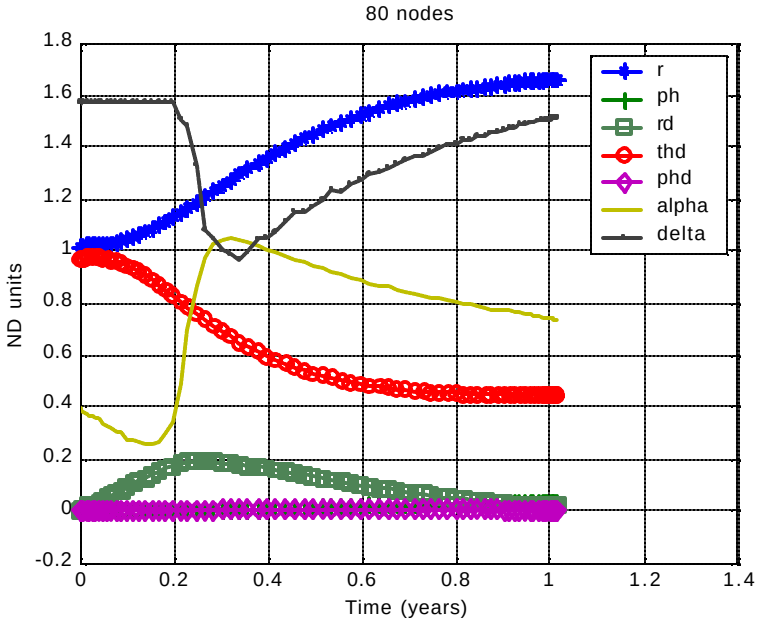

Figure 47 EM Rendezvous with Elliptic, Inclined Orbits.

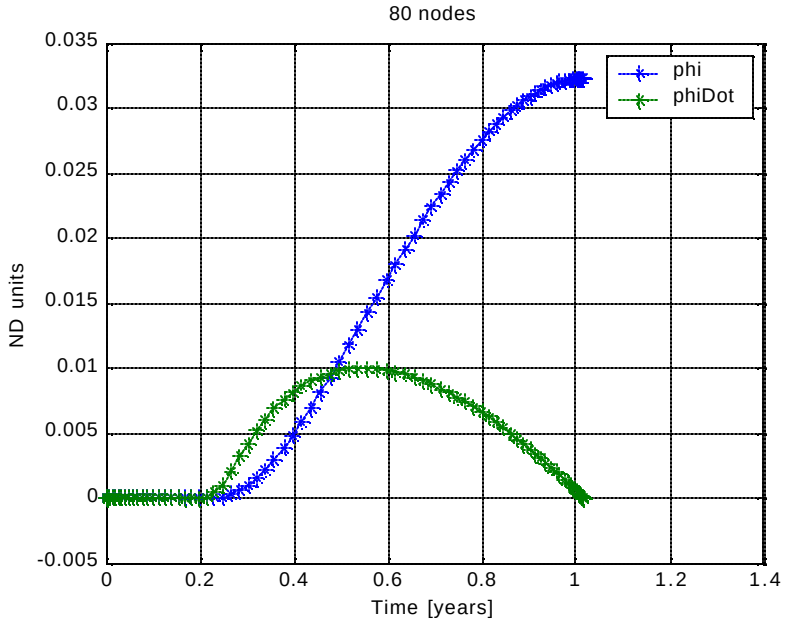

Figure 48 Optimal Profile for f and w_f States in the 3D Rendezvous Mission.

Turning to the costate analysis, it is evident that as in the 2-D solutions there is a critical time during the maneuver at which all costates are at a local maximum or minimum or zero (Figure 49). It is beyond the scope of this thesis as to why this behavior occurs, however at this point in time (about $1/4^{th}$ into the mission time) the control angles make large adjustments to achieve the state changes previously described. When inserted into the control laws derived in equations (29) and (31), the costates and states will produce derived cone and clock angle controls. A comparison of costate-derived controls and DIDO controls is shown in Figure 50. The optimal control does not rotate the cone angle to the extremes that the coplanar model control did. The controls here only pressed the cone angle between 18 and 60 degrees versus between 10 and 70 degrees. The optimal cone angle has to account for the clock angle as well in this 3 DOF model.

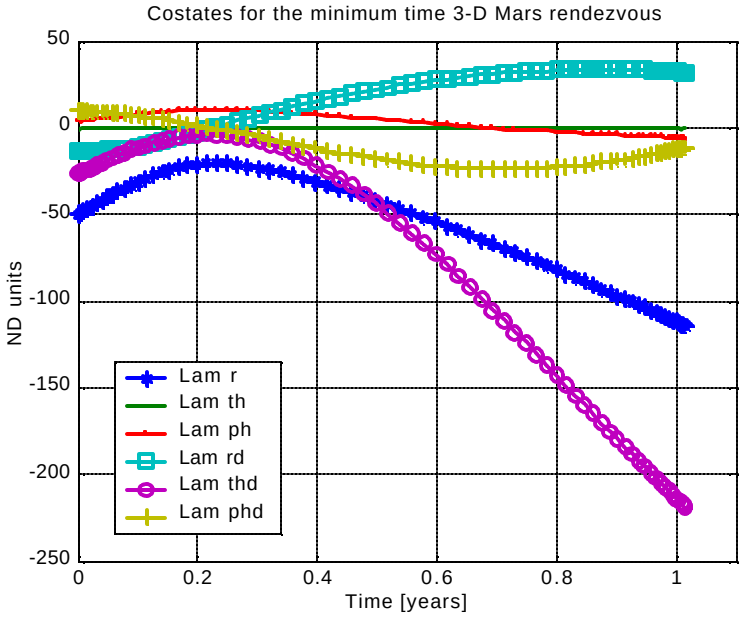

Figure 49 Costate History for Minimum Time Rendezvous with 3D Orbits Model.

Checking feasibility, the propagated controls were observed to match closely with the DIDO output (Figure 51). For optimality, the Hamiltonian is constant in Figure 52.

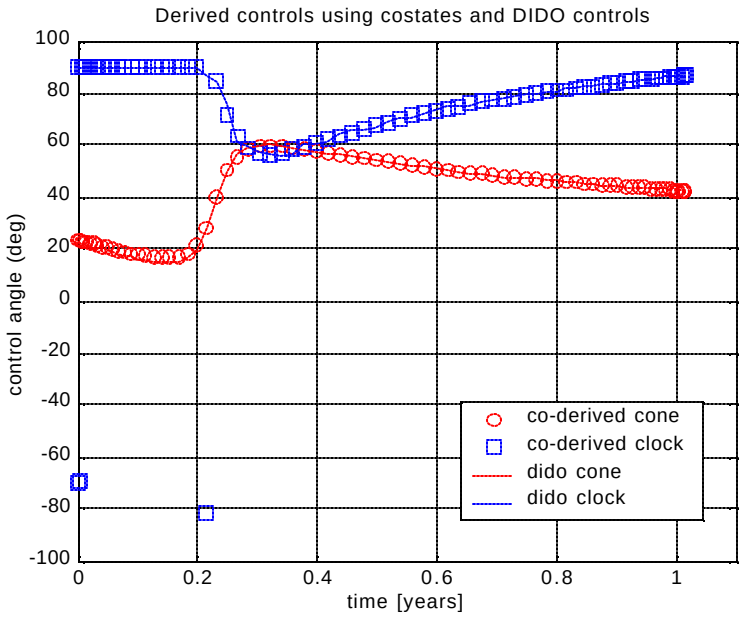

Figure 50 Control History for 3D Rendezvous. Displayed are DIDO controls (markers) and the controls derived from the tangent steering law with the states and costates as inputs (lines).

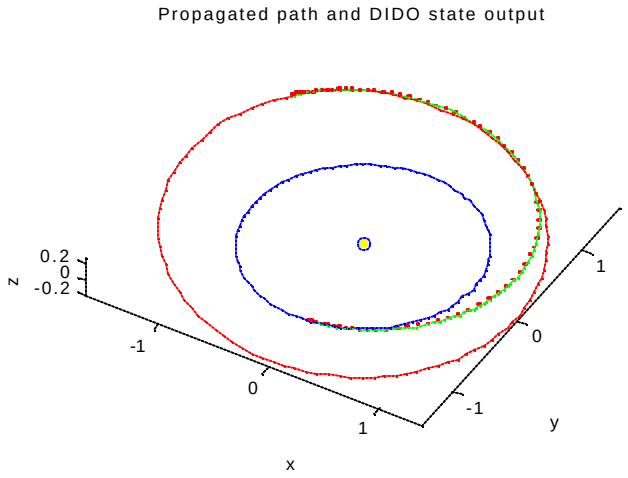

Figure 51 Propagated Path of 3D Rendezvous (line) with DIDO Output (dots). Mars inclination is exaggerated for display purposes.

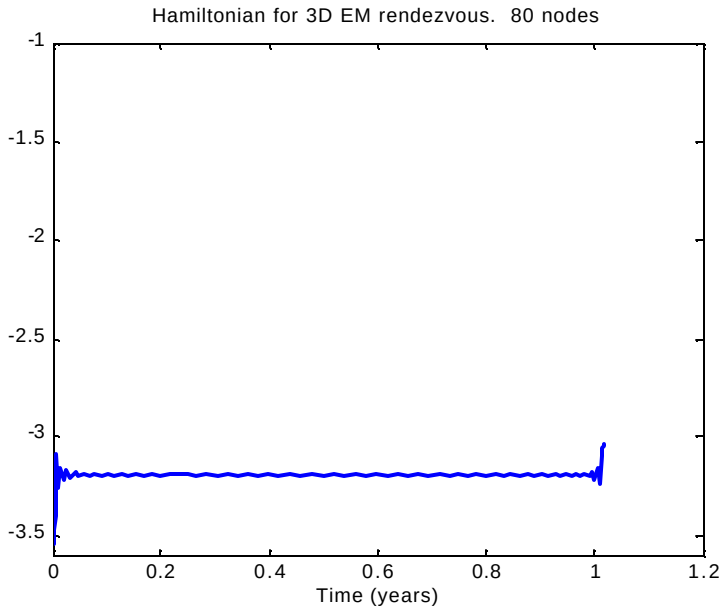

Figure 52 Hamiltonian of 3D EM Rendezvous Problem

c. Double Rendezvous

To optimize the Earth-Mars-Earth (EME) double rendezvous, we add to the rendezvous event conditions a return trip to Earth. The 2-D elliptic coplanar double-rendezvous exhibited symmetry in states and controls on the outbound and inbound legs (Figure 31). Now using inclined ellipses with their corresponding positions and velocity vector fields as target manifolds, we seek insights into the double-rendezvous with this higher fidelity model.

The event conditions now include an interior event corresponding to a rendezvous with Mars and a final event corresponding to a rendezvous with Earth. To ensure Earth is actually present when the sail intersects its orbit manifold, we constrain the time of flight using Kepler's equation. This last non-linear constraint replaces the corresponding one in the 2-D model that was listed in Table 4. Fortunately, the non-linear equation does not need to be solved explicitly as it simply serves as a constraint to the trajectory. In this way, the time of flight is coupled with the eccentric anomaly that the Earth traverses, E. Earth's angular displacement traversed is constrained by the

flight time of the spacecraft. As long as Earth's true anomaly change is the same displacement as the spacecraft angular displacement plus a multiple of 2π, phasing will have been achieved.

In order to rendezvous with a planet, we need to describe the position and velocity of both spacecraft and planet in 3-D (N_D=3) for all 4 events (N_E=4).

Spacecraft 3-D position at events in E-frame Cartesian coordinates are expressed as

$$\mathbf{R}_{sE} = \begin{pmatrix} x_0 & x_i^- & x_i^+ & x_f \\ y_0 & y_i^- & y_i^+ & y_f \\ z_0 & z_i^- & z_i^+ & z_f \end{pmatrix}$$

and likewise spacecraft 3-D velocity at events in LVLH.

$$\mathbf{V}_s = \begin{pmatrix} v_{r0} & v_{ri}^- & v_{ri}^+ & v_{rf} \\ v_{q0} & v_{qi}^- & v_{qi}^+ & v_{qf} \\ v_{f0} & v_{fi}^- & v_{fi}^+ & v_{ff} \end{pmatrix}$$

Transforming the sail velocity vector from LVLH into in E-frame, we use the transformation matrix.

$$\mathbf{B} = \underset{2\to 3}{[B]} = \left[R_3(-q)\right]\left[R_2(f)\right]$$

So spacecraft velocity in E-frame Cartesian coordinates is

$$\mathbf{V_{sE}} = \mathbf{B}\mathbf{V_s} \quad V_{sE} \in \mathbb{R}^{N_D \times N_E}$$

Because the craft is returning to Earth, it must be properly phased with Earth to ensure the planet is there when the spacecraft arrives. This constraint makes an Earth orbital motion model necessary accounting for its position in space and time. The following equations are used to constrain initial and final Earth events to ellipses in three dimensions. All state vector multiplication and division operators are element-wise.

First, we define the Earth initial and final angular positions in the perifocal frame.

$$\mathbf{q_e} = \begin{bmatrix} q_{e0} & q_{ef} \end{bmatrix} = \begin{bmatrix} q_0 & q_f \end{bmatrix} - w_e \begin{bmatrix} n_e & n_e \end{bmatrix},$$

Notice that $\mathbf{q_e} \in \mathbb{R}^2$ where $\begin{bmatrix} q_0 & q_f \end{bmatrix}$ are the initial and final angular spacecraft positions in E-frame which are coincident with Earth's initial and final position. The following equations relate Earth's eccentric anomaly to its true anomaly. Eccentric anomaly is expressed as a vec tor with each element corresponding to a knot, or end condition. The purpose of expressing it in this manner is to be able to handle multiple cycles with multiple encounters ($N_E > 1$) using a single variable.

The initial and final cosin e of Earth's eccentric anomaly are contained in the vector **cosE**. *The vector operations are performed in an "element-wise" fashion where the cosine of the vector* $\mathbf{q}_e$ *is the cosine of each individual element of v ector* $\mathbf{q}_e$.

$$\mathbf{cosE} = \frac{e_e + \cos(\mathbf{q_e})}{1 + e_e \cos(\mathbf{q_e})}, \quad \mathbf{cosE} \in \mathbb{R}^{N_E}$$

Earth's radial position as a function of **cosE** is given by

$$\mathbf{r_e} = a_e (1 - e_e \mathbf{cosE}) \quad , \ \mathbf{r_e} \in \mathbb{R}^{N_E}$$

where $\mathbf{r}_e = [r_{e0}, r_{ef}]^T$ so that vector **sinE** may be expressed as

$$\mathbf{sinE} = \frac{\mathbf{r}_e \sin(\mathbf{q_e})}{a_e \sqrt{1 - e_e^2}} \quad , \ \mathbf{r_e} \in \mathbb{R}^{N_E}$$

Thus, Earth's initial and final eccentric anomaly are contained in the vector **E**.

$$\mathbf{E} = \tan^{-1}\left(\frac{\mathbf{sinE}}{\mathbf{cosE}}\right) \quad , \ \mathbf{E} \in \mathbb{R}^{N_E} \qquad \text{Earth's eccentric anomaly}$$

The next set of equations is used to define Earth's position based on the initial and final time.

$$n = \sqrt{\frac{m}{a_e^3}} \qquad \text{Earth's mean motion}$$

$$M_0 = E_0 + e_e \sin E_0 \qquad \text{Initial mean anomaly}$$

$$M_f = n_e t_{TRANSIT} \qquad \text{Final mean anomaly}$$

Now we define the initial and final Earth position event matrix in heliocentric Cartesian coordinates, E-frame where.

$$\boxed{\mathbf{R_{eE}} = \begin{pmatrix} r_{e0}\cos(q_0) & r_{ef}\cos(q_f) \\ r_{e0}\sin(q_0) & r_{ef}\sin(q_f) \\ 0 & 0 \end{pmatrix}}$$

The speed of Earth on its elliptical path around the sun is

$$\mathbf{V_e} = \sqrt{m\left(\frac{2}{|\mathbf{r}_e|} - \frac{1}{a_E}\right)}\ , \quad \mathbf{V}_e \in \mathbb{R}^{N_E}$$

with velocity components expressed as

$$\mathbf{cos ß_e} = \frac{h_e}{|\mathbf{r}_e| \otimes \mathbf{V}_e}, \ \mathbf{cos ß_e} \in \mathbb{R}^{N_E}$$

$$\mathbf{sin ß_e} = \frac{m}{h_e \mathbf{V_e}} e_E \sin(\mathrm{q_e}), \ \mathbf{sin ß_e} \in \mathbb{R}^{N_E}$$

where $\mathbf{cos ß_e}$ and $\mathbf{sin ß_e}$ are the cosine and sine of Earth's flight path angle at the event knots. Earth's velocity in radial and transverse components in the perifocal frame are expressed in the following equations where the symbol $\otimes$ is used to emphasize that this is an "element-wise" multiplication such that each element in one vector is multiplied by the corresponding element in the other vector.

$$\mathbf{u_e} = \mathbf{V_e} \otimes \mathbf{sin ß_e} \qquad , \ \mathbf{u_e} \in \mathbb{R}^{N_E}$$

$$\mathbf{v_e} = \mathbf{V_e} \otimes \mathbf{cos ß_e} \qquad , \ \mathbf{v_e} \in \mathbb{R}^{N_E}$$

$$\mathbf{V_{eCART}} = \left[R_3(-\mathrm{q_e})\right] \begin{pmatrix} \mathbf{u_e} \\ \mathbf{v_e} \\ 0 \end{pmatrix} \qquad , \ \mathbf{V_{eCART}} \in \mathbb{R}^{N_D \times N_E}$$

$$\boxed{\mathbf{V_{eE}} = \mathbf{A_e}\mathbf{V_{eCART}}} \qquad , \ \mathbf{V_{eE}} \in \mathbb{R}^{N_D \times N_E}$$

where $\mathbf{A_e}$ is a matrix performing a 3-1-3 transformation from Earth perifocal to Cartesian heliocentric inertial (E-frame). See Appendix D for the derivation.

Having defined all the necessary quantities, it is now a simple matter to set boundary conditions. Letting the spacecraft initial and final event states be R_{sEnd} and V_{sEnd}, we have

$$\mathbf{R_{sEnd}} = \begin{pmatrix} x_0 & x_f \\ y_0 & y_f \\ z_0 & z_f \end{pmatrix} \quad , \quad \mathbf{V_{sEnd}} = \begin{pmatrix} V_{x_0} & V_{x_f} \\ V_{y_0} & V_{y_f} \\ V_{z_0} & V_{z_f} \end{pmatrix}$$

and intermediate event conditions are

$$\mathbf{R}_i = \begin{pmatrix} x_i^- & x_i^+ \\ y_i^- & y_i^+ \\ z_i^- & z_i^+ \end{pmatrix} \ , \ \mathbf{V}_i = \begin{pmatrix} V_x^- & V_x^+ \\ V_y^- & V_y^+ \\ V_z^- & V_z^+ \end{pmatrix}$$

so the boundary conditions at the earth encounter events become

$$\boxed{\mathbf{R_{End}} - \mathbf{R_{eE}} = \mathbf{0}}$$

and

$$\boxed{\mathbf{V_{End}} - \mathbf{V_{eE}} = \mathbf{0}}$$

We can establish the Mars intermediate events in a similar fashion.

$$\boxed{\mathbf{R}_i - \mathbf{R}_{mE} = \mathbf{0}}$$

and

$$\boxed{\mathbf{V}_i - \mathbf{V}_{mE} = \mathbf{0}}$$

One last event constraint ensures proper phasing with Earth. This is accomplished by forcing Kepler's equation as a constraint.

$$\boxed{E - e_E \sin E - M_0 - M_f = 0}$$

Earth mean anomaly and eccentric anomaly are represented by M and E respectively. The event constraints for the 3D EME rendezvous mission are summarized in the table below.

Event	***Double rendezvous constraints [lower,upper]***
$\mathbf{R}_{End}$	$[\mathbf{R}_{eE}, \mathbf{R}_{eE}]$
$\mathbf{V}_{End}$	$[\mathbf{V}_{eE}, \mathbf{V}_{eE}]$
$\mathbf{R}_i$	$[\mathbf{R}_{mE}, \mathbf{R}_{mE}]$
$\mathbf{V}_i$	$[\mathbf{V}_{mE}, \mathbf{V}_{mE}]$
$E - e_E \sin E - M_0 - M_f$	[0,0]

Table 11 EME 3D Double Rendezvous Event Constraints

States and controls for the 3 DOF model are bounded by

$$\mathbf{x} \in S \subset \mathbb{R}^6$$
$$\mathbf{u} \in U \subset \mathbb{R}^2$$

Where the set S is chosen to avoid singularities in the dynamics and the set U was chosen to avoid *duplicity* in the controls. Because it was desired to obtain distinct (α, δ) pairs (every sail normal vector orientation is represented by only one (α, δ) pair), the bounds were such that $0 \le \alpha \le \frac{\pi}{2}$ and $-\pi \le \delta < \pi$. It was recognized that the sin and cos components of the normal vector could have been used as the controls with proper path constraints, however the numerical solutions took longer to complete, so the cone and clock angles were retained as the controls.

Results and Analysis

Once more, Mars aphelion played the role of the optimal rendezvous point at the intermediate event for the round trip. Modeling Mars' small inclination with the

ecliptic produced a solution that looked very similar to the 2D elliptic when case when looking from the top view (compare Figure 53 with Figure 31). The mission time, however, was slightly longer than the 2D case taking 2.42 years instead of 2.3 years since the spacecraft must change planes twice. The mission time of the 3D model is the same as that of the 2D circular orbits model except that there is no longer the precise symmetry of states and controls exhibited by the other model. In three dimensions, the spacecraft takes 1.03 years on the outbound leg and 1.39 on the inbound (2D circular coplanar model required 1.205 years/leg). One of the amazing features of the path is that the spacecraft sails out of both Mars and Earth orbital planes for a significant portion of the return trip to Earth rendezvous (Figure 54). This large departure from the planetary planes is accomplished in order to reach Earth orbit in phase with Earth to an accuracy 2.64E-4 radians (~39E3 km from the center of the Earth). Testing the feasibility, the control profile was entered into the propagator producing the result in Figure 55.

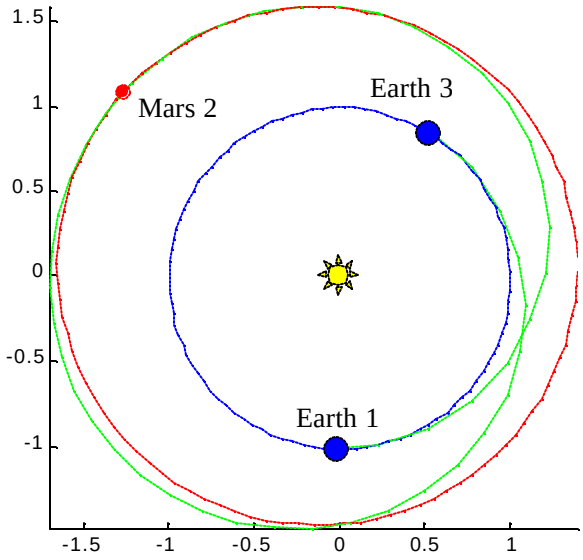

Figure 53 3D EME Double Rendezvous Path (top view)

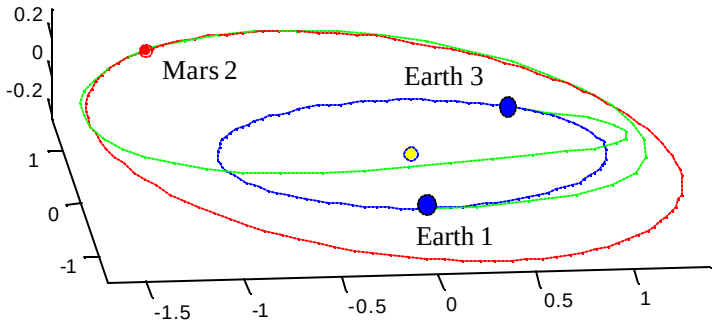

Figure 54 3D EME Double Rendezvous Path (oblique view)

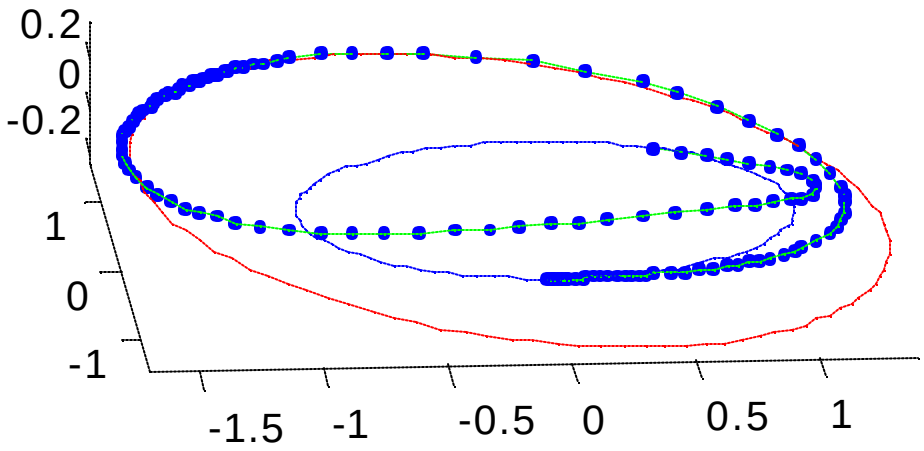

Figure 55 Propagated Path (line) and DIDO States (dots) for 3D EME Double Rendezvous

2. Earth-Mars Cycler Set-up

Formulating the 3-D elliptic orbit cycler model into an OCP is a formidable task because cyclic end conditions are no longer imposed every EME round trip. Many round trips occur before Mars and Earth positions are exactly the same as they were at the start of the trajectory. The events as structured in the previous section are set up to handle multiple passes for a "cycle". Since meeting synodic conditions with each round trip is no longer required, a possible OCP formulation would be to seek the minimum time to achieve a given number of EM passes. Using the framework outlined in this thesis, this would entail performing successive solutions to increasing number of EM visits (and thus number of knots) using intercept end conditions until initial state conditions are achieved at the final time.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND FUTURE WORK

A. CONCLUSIONS

Solar sails provide a viable means of propulsion for an Earth-Mars cycler. For a solar sail with lightness number $b = .17$, an orbit can be maintained with rapid revisit times between destination planets and slow V_∞s at each planet, both desirable cycler characteristics. Furthermore, the cycler lifetime is only dependent on sail degradation in the space environment since no propellant is depleted making solar sail cyclers attractive in comparison to conventional low thrust and chemical propellant cyclers.

The framework in this thesis outlines a method of state and control optimization for both 2-D and 3-D cycler models. This framework can be applied to other missions as well by changing the event conditions. With relatively simple dynamics and sail models, trajectory design could be accomplished by properly formulating desired events in the form of constraints. The pseudospectral method used within DIDO to solve the OCPs produces optimal controls and states that were verified as feasible (through propagation of initial conditions with a third party ODE solver) and optimal using the necessary conditions for optimality.

B. OPPORTUNITIES FOR FUTURE WORK

The following are some areas of interest for continuing work.

Earth-Venus-Mars cycler. Is a cycler possible with three target planets and if so, what does it look like? Is it possible to reduce the revisit times from a two-planet cycler? This problem would require the solution to include the optimal order of planetary encounters.

Use a solar sail to establish an optimal halo orbit. Many non-Keplarian orbits are possible with solar sails. Orbits around Lagrange points may be modeled and examined for missions such as early detection of geomagnetic storms on the sunward side of the Earth-Sun L1 point (ref 9 p. 231). Cycling orbits between L points could also be explored.

Implement two synodic period cyclers with solar sails. Reference 19 outlines a plan to make multiple passes of a target planet within a cycle (2 synodic periods). Planet-centered dynamics may play a role if trajectories spend significant time in a planet's SOI.

SEP/Solar Sail hybrid cyclers. Modeling a SEP/solar sail hybrid spacecraft would show when it is optimal to favor the SEP and when it is preferable to favor the sail for various cycler orbits (or any trajectory for that matter).

Variable Isp cycler. When is it optimal to use high thrust-low Isp and when is it best to use low thrust-high Isp in a cycler trajectory? There are trades to examine with weighted cost functions between fuel mass and time.

Make smoother controls near hard knots. Changing the controls to include "inertia" assisted in making controls somewhat smoother near knots, however for 3D problems it was important to implement constraints to ensure that arbitrary sail orientations resulted from unique cone angle/clock angle pairs. Implementing the sin and cos components of these angles as controls (with a path constraint) slowed the algorithm down and was not implemented, however this may have alleviated some of the jumpy control outputs.

Improve gravity assist model. The matched asymptotes model as presented in this thesis is accurate for the problems posed, however more accurate solutions to some problems may call for a better gravity assist model. Possibilities are to either change dynamics inside SOI or use a jump in time and theta at the interior knot in addition to velocity.

Improve sail model. A higher fidelity model of the sail is available that includes billowing and diffuse reflection (see ref 9 pp.51-54). Once the dynamics and events models have been thoroughly tested, there are some design trades that could be explored as far as the sail itself.

Change to dynamic equations that are better suited for cyclers. The discussion on coordinate systems outlined some benefits and pitfalls of using certain coordinates in cycler trajectories. Perhaps other coordinates than those presented here can be better implemented for optimal cyclers (e.g. equinoctial).

Solar sail stability. Solar sails have been used in spacecraft to dump accumulated momentum in reaction wheels due to torque perturbations. A possible OCP would be to minimize the time to dump momentum in reaction wheels for a given earth-centered orbit.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. APPLICATION OF THE MINIMUM PRINCIPLE[10,17]

The first step in solving an optimal control problem is to construct the scalar Hamiltonian function, H,

$$H(\mathbf{x},\mathbf{u},t,\mathrm{l}) = F(\mathbf{x},\mathbf{u},t) + \mathrm{l}(t)^{\mathrm{T}}\mathbf{f}(\mathbf{x},\mathbf{u},t)$$

where $\mathbf{f}(\mathbf{x},\mathbf{u},t)$ are the dynamic constraints on the system, and $\mathrm{l}(t)$ are the Lagrange multipliers called costates. According to Pontryagin's Minimum Principle, the optimal control is obtained by solving the following problem at each instant in time.

Minimize H with respect to $\mathbf{u}$, with $\mathbf{u} \in U$ where U is the set of all allowable controls. To solve this problem, the Lagrangian of the Hamiltonian is written as:

$$L(\mathbf{x},\mathbf{u},t,\mathrm{l},\boldsymbol{\mu}_g) = H(\mathbf{x},\mathbf{u},t,\mathrm{l}) + \boldsymbol{\mu}_g(t)^{\mathrm{T}}\mathbf{g}(\mathbf{x},\mathbf{u},t)$$

where $\mathbf{g}(\mathbf{x},\mathbf{u},t)$ are the path constraints and $\boldsymbol{\mu}_g(t)$ are the associated path covectors. The path constraints include all trajectory path constraints as well as any state and control bounds. Applying the Karush-Kuhn-Tucker (KKT) theorem to the Lagrangian results in a set of first order necessary conditions and provides a means to demonstrate the optimality of a solution.

(32)
$$\frac{\partial L}{\partial \mathbf{u}} = \mathbf{0}$$

$$\boldsymbol{\mu}_g(t)^{\mathrm{T}}\mathbf{g}(\mathbf{x},\mathbf{u},t) = \mathbf{0}$$

with

$$\boldsymbol{\mu}_g = \begin{cases} \leq 0 \\ \geq 0 \\ = 0 \\ any \end{cases} \quad if \quad \begin{cases} g_l = g(\mathbf{x},\mathbf{u},\tau) \\ g(\mathbf{x},\mathbf{u},\tau) = g_u \\ g_l < g(\mathbf{x},\mathbf{u},\tau) < g_u \\ g_l = g_u \end{cases}$$

The third case above describes a special condition when the constraints in $\mathbf{g}$ are interior or non-binding,

$$\mathbf{g}_l < \mathbf{g}\left(\mathbf{x}, \mathbf{u}, t\right) < \mathbf{g}_u$$

For these cases, the multipliers, $\boldsymbol{\mu}_g = \mathbf{0}$ and equation (32) simplifies to

$$\frac{\partial L}{\partial \mathbf{u}} = \frac{\partial H}{\partial \mathbf{u}} = \mathbf{0}$$

A necessary second order condition for optimality when $\boldsymbol{\mu}_g = \mathbf{0}$ is that

$$\frac{\partial^2 H}{\partial \mathbf{u}^2} = \frac{\partial^2 L}{\partial \mathbf{u}^2} \geq \mathbf{0} \tag{33}$$

APPENDIX B. DERIVATION OF THE 2-D EQUATIONS OF MOTION

Lagrangian mechanics are used to achieve the 2-D equations of motion using polar coordinates. Velocity variables are then written into a more familiar form using the radial and transverse velocity components. Figure 1 shows the states and coordinate convention used in the derivation. In general form Lagrange's equation is written as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}}\right)-\frac{\partial L}{\partial \mathbf{x}}=\mathbf{Q}_{nc}{}^{T}$$

where L is the Lagrangian, $\mathbf{x}$ is the generalized coordinate vector, and $\mathbf{Q}_{nc}$ is the generalized non-conservative force vector[18].

We start with writing the specific kinetic energy as

$$T=\frac{1}{2}\mathbf{v}\cdot\mathbf{v}=\frac{1}{2}(\dot{r}^2+r^2\dot{q}^2)$$

Potential energy due to gravity is

$$V=-\frac{GM}{r}=-\frac{m}{r}$$

The Lagrangian may be written as

$$L=T-V=\frac{1}{2}(\dot{r}^2+r^2\dot{q}^2)+\frac{m}{r}$$

The non-conservative generalized force is due to the solar radiation pressure (SRP) on the solar sail. Acceleration magnitude is given by

$$\frac{T}{m}=\frac{bm}{r_0^{\,2}}\cos^2 a$$

Thus the acceleration in the radial direction is

$$a_r=\frac{T}{m}\cos a=\frac{bm}{r_0^{\,2}}\cos^3 a$$

and the applied acceleration in the along-track direction is

$$a_q = \frac{T}{m}\sin a = \frac{bm}{r_0^{\,2}}\cos^2 a \,\sin a$$

Writing the Lagrangian equations of motion, we obtain

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}}\right) = \ddot{r} \qquad \frac{\partial L}{\partial r} = r\dot{q}^2 - \frac{m}{r^2}$$

(34) $$\ddot{r} - r\dot{q}^2 + \frac{m}{r^2} = a_r$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 2r\dot{r}\dot{q} + r^2\ddot{q} \qquad \frac{\partial L}{\partial q} = 0$$

(35) $$2r\dot{r}\dot{q} + r^2\ddot{q} = a_q r$$

where $a_q r$ is a generalized specific torque.

Equation (35) reduces to the conservation of angular momentum when $a_q = 0$.

Now to convert the $\dot{r}$s and $\dot{q}$s into u and v velocity components, we use the following relations.

Transverse velocity is

(36) $$v = r\dot{q}$$

and transverse acceleration is

(37) $$\dot{v} = r\ddot{q} + \dot{r}\dot{q}$$

Using the relation in equation (36) with equation (34) we write the equation of motion for $\dot{u}$.

(38) $$\boxed{\dot{u} = \frac{v^2}{r} - \frac{m}{r^2} + \frac{bm}{r_0^{\,2}}\cos^3 a}$$

Rewriting equation (35) and factoring an r yields

(39) $$r\left(r\ddot{q} + \dot{r}\dot{q} + \dot{r}\dot{q}\right) = a_q r$$

Recall that $u = \dot{r}$ therefore $uv = \dot{r}r\dot{q}$. Using (37) with (39) we obtain the equation of motion for $\dot{v}$.

$$r\dot{v} + r\dot{r}\dot{q} = a_q r$$

$$\dot{v} = -\frac{uv}{r} + a_q$$

(40)
$$\boxed{\dot{v} = -\frac{uv}{r} + \frac{bm}{r_0^2}\cos^2 a \sin a}$$

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. DERIVATION OF THE 3-D EQUATIONS OF MOTION

As with the 2-D case, we use the Lagrangian approach to acquire the 3-D equations of motion.

Figure 56 shows the RSW spherical coordinate system (ref 4 p. 42) and convention used to represent the states.

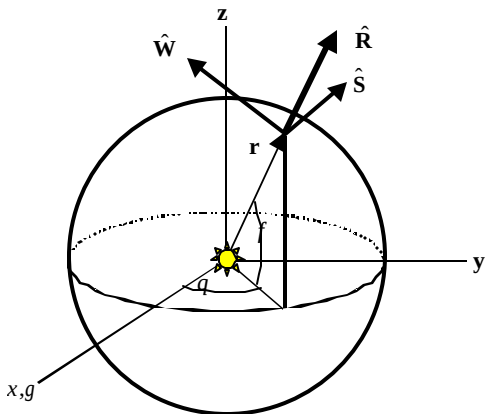

Figure 56 RSW Coordinate System

In general form Lagrange's equation is written as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}}\right)-\frac{\partial L}{\partial \mathbf{x}}=\mathbf{Q}_{\mathbf{nc}}{}^{T} \tag{41}$$

where L is the Lagrangian, $\mathbf{x}$ is the generalized coordinate vector, and Q_{nc} is the generalized non-conservative force vector[18]. The generalized non-conservative force for this case is the solar radiation pressure (SRP) on the solar sail resolved in radial ($\hat{\mathbf{R}}$), along-track ($\hat{\mathbf{S}}$), and cross-track ($\hat{\mathbf{W}}$) directions.

We start by writing the specific kinetic energy as

$$T=\frac{1}{2}\mathbf{v}\cdot\mathbf{v}$$

where the velocity $\mathbf{v}$ is

$$\mathbf{v}=\dot{r}\hat{\mathbf{R}}+r\dot{q}\cos f\hat{\mathbf{S}}+r\dot{f}\hat{\mathbf{W}}$$

Specific potential energy is

$$V=-\frac{m}{r}$$

making the Lagrangian

$$L=T-V=\frac{1}{2}\left[\dot{r}^{2}+\left(r\dot{q}\cos f\right)^{2}+\left(r\dot{f}\right)^{2}\right]+\frac{m}{r}$$

The external acceleration due to the SRP on the sail has a magnitude of

$$a=\frac{T}{m}=\frac{bm}{r^{2}}\cos^{2}a$$

and have components in the radial, along-track and cross-track directions, written as

$$a\cos a\,\hat{\mathbf{R}}+a\sin a\,\sin d\,\hat{\mathbf{S}}+a\sin a\,\cos d\,\hat{\mathbf{W}}$$

Having established the necessary quantities, we proceed to derive the Lagrangian equations of motion. For the r state, we have the following.

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}}\right)=\ddot{r}\qquad\frac{\partial L}{\partial r}=r\dot{q}^{2}\cos^{2}f+r\dot{f}^{2}-\frac{m}{r^{2}}$$

(42) $$\boxed{\ddot{r} - r\dot{q}^2 \cos^2 f - r\dot{f}^2 + \frac{m}{r^2} = a\cos a}$$

Turning to the θ state we form the quantities

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = r^2\dot{q}\left(-2\dot{f}\cos f \sin f\right) + \cos^2 f\left(r^2\ddot{q} + 2r\dot{r}\dot{q}\right) \qquad \frac{\partial L}{\partial q} = 0$$

The generalized force is a torque about the $\hat{\mathbf{W}}$ axis

$$Q_w = a\sin a \sin d\left(r\cos f\right)$$

Lagrange's equation (41) becomes

$$r^2\dot{q}\left(-2\dot{f}\cos f \sin f\right) + \cos^2 f\left(r^2\ddot{q} + 2r\dot{r}\dot{q}\right) = a\sin a \sin d\left(r\cos f\right)$$

Dividing by $r\cos f$ results in

$$r\dot{q}\left(-2\sin f\dot{f}\right) + \cos f\, r\ddot{q} + 2\dot{r}\dot{q} = a\sin a \sin d$$

Finally, we rearrange terms to get the equation of motion for $\ddot{q}$.

(43) $$\boxed{\ddot{q} = \frac{1}{r\cos f}\left[2\dot{q}\left(r\dot{f}\sin f - \dot{r}\cos f\right) + a\sin a \sin d\right]}$$

Note that we need to be cautious at the singularities where r=0 and cos ϕ=0.

Now for the f state we write

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{f}}\right) = 2r\dot{r}\dot{f} + r^2\ddot{f} \qquad \frac{\partial L}{\partial f} = -r^2\dot{q}^2\cos f \sin f$$

The generalized force in the f sense direction (torque in this case) is given by

$$Q_f = ar\sin a \cos d$$

Substituting into equation (41) yields

$$2r\dot{r}\dot{f} + r^2\ddot{f} + r^2\dot{q}^2\cos f \sin f = ar\sin a \cos d$$

Dividing by r and rearranging gives the equation of motion for $\ddot{f}$.

(44) $$\boxed{\ddot{f} = \frac{1}{r}\left(-2\dot{r}\dot{f} + a\sin a\,\cos d\right) - \dot{q}^2 \sin f\,\cos f}$$

APPENDIX D. TRANSFORMATION INTO HELIOCENTRIC - ECLIPTIC COORDINATES (E-FRAME)

The matrix derivations below are used to transform the spacecraft state and event conditions into a common coordinate system – heliocentric ecliptic Cartesian coordinates, called the E-frame. *Spacecraft* states at the events (position and velocity) are given in *spherical* coordinates while the target planet manifolds (elliptical paths with velocity related to position) are most easily expressed in a perifocal frame.

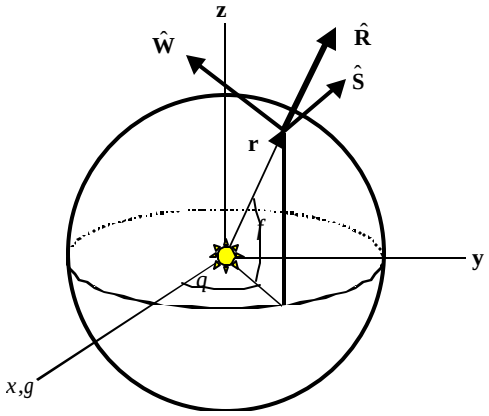

Figure 57 Spherical coordinate system

SPACECRAFT

Referring to the Figure 57, we transform the spherical position coordinates into Cartesian coordinates as follows. It is worth noting that some texts (e.g. ref 18) choose the ϕ coordinate as the complement angle from the one depicted above, so it is important to be careful in deriving the transformation matrices.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_E = \begin{bmatrix} r\cos f\cos q \\ r\cos f\sin q \\ r\sin f \end{bmatrix}$$

The spacecraft velocity state has elements given as $\dot{r}, \dot{q}, \dot{f}$. In the parlance of DIDO, these spacecraft states are the "primal.states". First we need the velocity vector of the spacecraft in v_r, v_q, v_f components where the subscripts indicate which sense direction the velocity vectors point. This transformation appears as

$$\begin{bmatrix} v_r \\ v_q \\ v_f \end{bmatrix} = \begin{bmatrix} \dot{r} \\ r\dot{q}\cos f \\ r\dot{f} \end{bmatrix}$$

or in our notation

$$\begin{bmatrix} v_r \\ v_q \\ v_f \end{bmatrix} = \begin{bmatrix} v_r \\ rw_q \cos f \\ rw_f \end{bmatrix} \tag{45}$$

These coordinates may now be transformed into the E-frame using a 3-rotation and 2-rotation matrix.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}_E = \left[\mathbf{R}_3(-q)\right]\left[\mathbf{R}_2(f)\right]\begin{bmatrix} v_r \\ v_q \\ v_f \end{bmatrix}$$

We choose to call the matrix $\left[\mathbf{R}_3(-q)\right]\left[\mathbf{R}_2(f)\right]$ the **B** matrix.

$$\boxed{\mathbf{B} = \begin{bmatrix} \cos q \cos f & -\sin q & \cos q \sin f \\ \sin q \cos f & \cos q & \sin q \sin f \\ -\sin f & 0 & \cos f \end{bmatrix}}$$

Using transformations given in boxes, we can transform the position and velocity of the spacecraft in spherical coordinates (primal.states) into the E-frame coordinate frame.

PLANETS

To define the manifolds of Earth and Mars we need their positions, defining an ellipse, and equations for their velocities at any point on the ellipse. Time is imposed as a non-linear constraint in the event conditions without need for transformation, so it will not be addressed here. An ellipse may be represented easily in either polar or Cartesian form, so we choose the polar form since most astrodynamic textbooks use this form. Positions and velocities in the perifocal frame (Figure 58) are related by the following equations.

$$r = \frac{p}{1 + e\cos q_p} \text{ and } V = \sqrt{m\left(\frac{2}{r} - \frac{1}{a}\right)}$$

The q_p used in these equations is the angular positions of the *planet* from their perihelion at a particular event. The relation between the planet true anomaly and the spacecraft angular position is

$$q_p \approx q - \left(w_p + \Omega_p\right)$$

for small inclinations (ref 4). For larger inclinations another transformation is required.

Taking the locus of planet positions, the ellipse, and transforming in to Cartesian coordinates we write

$$\begin{aligned} x_p &= r\cos q_p \\ y_p &= r\sin q_p \\ z_p &= 0 \end{aligned}$$

Transforming into the E-frame requires a 3-1-3 rotation matrix defined using the longitude of ascending node Ω, inclination i, and argument of periapsis w (ref 13).

This matrix that transforms Cartesian coordinates from the perifocal frame to the E-frame is defined as

$$\underset{E\leftarrow p}{\mathbf{A}} = \left[\mathbf{R}_3(-w)\right]\left[\mathbf{R}_1(-i)\right]\left[\mathbf{R}_3(-\Omega)\right]$$

$$= \begin{bmatrix} \cos\Omega\cos w - \cos i\sin\Omega\sin w & -\cos i\sin\Omega\cos w - \cos\Omega\sin w & \sin i\sin w \\ \cos i\cos\Omega\sin w + \sin\Omega\cos w & \cos i\cos\Omega\cos w - \sin\Omega\sin w & -\sin i\cos\Omega \\ \sin i\sin w & \sin i\cos w & \cos i \end{bmatrix}$$

R denotes rotation matrices about the 1, 2, or 3 axis. See ref 13, p.80, ref 18, p.190, and ref 4, p. 151 for more details.

Now transforming the planetary velocity vectors from the perifoc al coordinates (Figure 58), u and v in figure to the E-frame. First, transform u and v into Cartesian coordinates in the perifocal frame.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}_{perifocal} = \left[\mathbf{R}_3 \left(-q_p \right) \right] \begin{bmatrix} u \\ v \\ 0 \end{bmatrix}$$

Finally, transforming the Cartesian velocity vector in the per ifocal plane into the E-frame, we use the **A** matrix again.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}_E = \mathbf{A} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}_{perifocal}$$

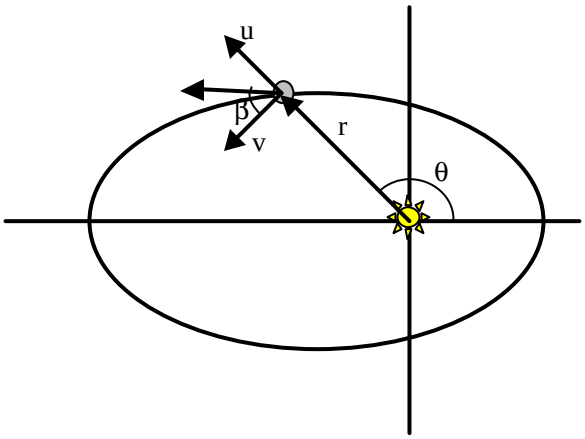

Figure 58 Planetary perifocal system

APPENDIX E. SOLAR SAIL CONTROL HODOGRAPH ANALYSIS

A hodograph is useful in determining convergence char acteristics of a solution using a particular set of controls in the dynamics. Figure 59 shows the hodograph, the mapping of the control, α, onto the $\dot{u}-\dot{v}$ plane. Because the thrust *magnitude* and resulting acceleration due to the sail is completely coupled with the sail angle, the controls map as closed curve, not a region. Equations (38) and (40) plotted for all feasible a values produces the hodograph. The curve is not convex in that a line drawn between any two points on the curve does not remain completely in the locus of points defined by the control mapping. There may be a relation of convexity and convergence speed.

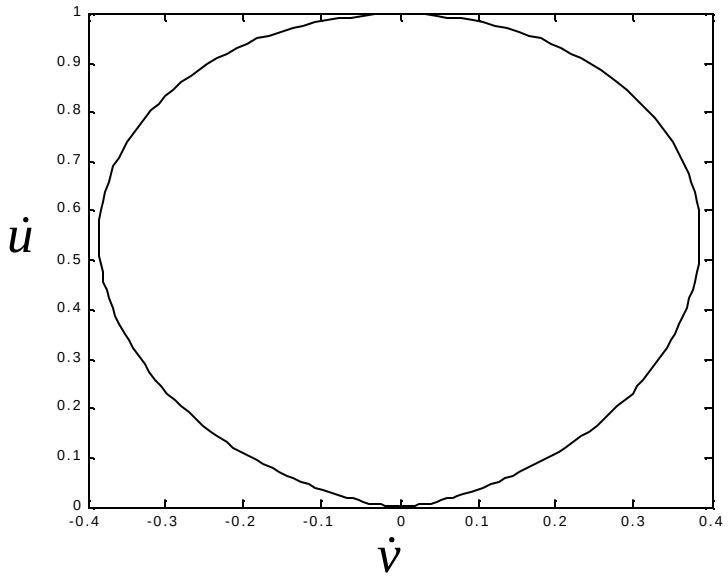

Figure 59 Hodograph for Cone Angle, a

Using a new control scheme by setting the control variable equal to $\dot{a}$ maps as a straight line, which is inherently convex using the above definition. Solutions do in fact converge faster for this case.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

1. Byrnes, D., Longuski, J., Aldrin, B., "Cycler Orbit Between Earth and Mars," *Journal of Spacecraft and Rockets*, Vol. 30, No. 3, May-June 1993, pp. 334-336.

2. Niehoff, J., "Pathways to Mars: New Trajectory Opportunities," American Astronautical Society, AAS Paper 86-172, July 1986.

3. Ross, I. M. and Fahroo, F., "User's Manual for DIDO 2001: A MATLAB Application Package for Dynamic Optimization," *NPS Technical Report AA-01-003*, Department of Aeronautics and Astronautics, Naval Postgraduate School, Monterey, CA, December 2001.

4. Vallado, D. A., *Fundamentals of Astrodynamics and Applications*, Second Edition, Space Technology Library, Microcosm Press, El Segundo, CA, 2001.

5. Strizzi, J., Ross, I. M. and Fahroo, F., "Towards Real-Time Computation of Optimal Controls for Nonlinear Systems," *AIAA Guidance, Navigation and Control Conference*, Monterey, CA 2002. Paper No. AIAA 2002-4945.

6. Yan, H., Fahroo, F., Ross, I., "Real-Time Computation of Neighboring Optimal Control Laws," Proceedings of the AIAA Guidance, Navigation and Control Conferences, August 2002.

7. Ross, I., Fahroo, F., "Convergence of Pseudospectral Approximations for Optimal Control Problems," Proceedings of the 2001 IEEE Conference on Decision and Control, December 2001.

8. Melton, R., "Comparison of Direct Optimization Methods Applied to Solar Sails Problems," *AIAA Astrodynamics Specialist Conference*, Monterey, CA, August 2002.

9. McInnes, C., *Solar Sailing*, Springer Publishing, ltd, Chichester, U.K., 1999

10. Bryson, A. E., *Dynamic Optimization*, Addison-Wesley Longman, Inc., 1999

11. Wright, J., *Space Sailing*, p. 2., Gordon and Breach, 1992.

12. Zhukov, A. N., Lebedev, V. N., "Variational Problem of Transfer Between Heliocentric Circular Orbits by Means of a Solar Sail", Translated from Kosmicheskie Issledovaniya, Vol. 2, No. 1, pp 45-50, January-February 1964.

13. Bate, R. R., Mueller, D. D., White, J. E., *Fundamentals of Astrodynamics*, Dover Publications, Inc., New York, 1971, pp. 399.

14. Aldrin, B., Byrnes, D., Jones, R., Davis, H., "Evolutionary Space Transportation Plan for Mars Cycling Concepts," American Institute of Aeronautics and Astronautics, 2001-4677

15. Prussing, J., Conway, B., Orbital Mechanics, Oxford University Press, New York, 1993.

16. Kaplan, M. H., *Modern Spacecraft Dynamics and Control*, John Wiley & Sons, New York, 1976.

17. Ross, I., AA4850 course notes, December 2001.

18. Baruh, H., *Analytical Dynamics*, McGraw-Hill Companies, Inc., 1999.

19. Byrnes, D., McConaghy, T., Longuski, J., "Analysis of Various Two Synodic Period Earth-Mars Cycler Trajectories," AAS Astrodynamics Specialist Conference Paper 2002-4423, 5-8 August 2002, *American Institute of Aeronautics and Astronautics.*

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Professor I. Mic hael Ross
   Naval Postgraduate School
   Monterey, California

4. Professor Fariba Fahroo
   Naval Postgraduate School
   Monterey, California

5. Dennis Byrnes
   Jet Propulsion Lab
   Pasadena, California

6. Jon Sims
   Jet Propulsion Lab
   Pasadena, California

7. Robert E. Stevens
   SPAWAR SFA
   Chantilly, Virginia